August 16, 2018

VIA ELECTRONIC SUBMISSION

The Honorable Andrew Wheeler
Acting Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Ave., N.W.
Washington, DC 20460

Attn: EPA–HQ–OA–2018–0259

Re: Comment of the Environmental Defense Fund on the Environmental Protection Agency's *Proposed Rule: Strengthening Transparency in Regulatory Science*, 83 Fed. Reg. 18768 (Apr. 30, 2018) ("Proposal")

Environmental Defense Fund ("EDF") submits the following comments on EPA's April 30, 2018 proposed rule, "Strengthening Transparency in Regulatory Science" (the "Proposal").[1] Representing over two million members and supporters, EDF applies science, economics, and the law to solve our most urgent public health and environmental problems. EDF regularly engages in policy advocacy, regulatory proceedings, and litigation to secure and defend protections for human health and the environment under the Clean Air Act ("CAA"), Toxic Substances Control Act ("TSCA"), and other statutes administered by EPA—protections that save lives, improve well-being, and provide a more vibrant economy for all Americans, including our members. EDF and our members therefore have a profound stake in ensuring that EPA actions are anchored in the best available science, and are not distorted by policies and practices that seek to unjustifiably limit EPA's use of science for the purpose of weakening health and environmental protections.

For the reasons explained below, the Proposal would violate EPA's substantive and procedural obligations, is arbitrary and capricious, and must be withdrawn. Indeed, the Proposal is the classic wolf in sheep's clothing. Cloaked in vague platitudes about scientific quality and promoting "transparency," the Proposal would establish a sweeping new regulatory requirement prohibiting EPA from considering public health studies for which underlying data cannot be made "publicly available in a manner sufficient for independent validation."[2] This requirement would bar EPA from considering many vital public health studies that are based on confidential patient information that cannot be legally or ethically disclosed, and have been rigorously vetted using time-tested approaches that are widely accepted in the scientific community. Nowhere does the Proposal document what deficiencies in existing EPA regulatory science it is trying to solve, much less why such draconian restrictions on the use of science would improve the quality of EPA decision-making.

This wolf's true nature, however, cannot be covered up: the Proposal is in fact directed at excluding the best available science demonstrating significant health and welfare effects from

[1] *Strengthening Transparency in Regulatory Science*, 83 Fed. Reg. 18,768 (Apr. 30, 2018).
[2] *Id.* at 18,773 (proposed 40 C.F.R. § 30.5).

agency decision-making in order to thwart the agency's ability to protect the public health and welfare. As our comments document, the Administration hastily concocted this Proposal as a way of unilaterally implementing failed legislative proposals backed by prominent opponents of accepted climate change science and patterned on proposals put forward by the tobacco industry in the 1990s. According to records obtained from EPA through the Freedom of Information Act when this Administration's own political staff discovered that earlier versions of the Proposal might also restrict industry-funded science supporting the registration of pesticides and other chemicals, it decided to "thread this one real tight!" to ensure that *only* those studies supporting public health regulations would be subject to this new "transparency" rule.[3]

Ultimately, this Proposal does not "strengthen science." EPA's Science Advisory Board ("SAB") and the scientific community were not even consulted in its development, and a host of scientific authorities—including members of the SAB, editors of the nation's leading scientific journals, the National Academies, and numerous scientific and medical organizations—have raised fundamental concerns about the Proposal. Rather than strengthen science, the Proposal grants the Administrator vague and manipulable authority to *censor* science that by any scientific definition is the best simply because it conflicts with this Administration's political goals. We urge EPA to abandon this deeply destructive and misguided Proposal.

Respectfully submitted,

Tomás Carbonell
Ben Levitan
Jennifer McPartland
Ryan O'Connell
Martha Roberts
Ananya Roy
Surbhi Sarang
Robert Stockman
Environmental Defense Fund

Keri Powell
Alexandra Teitz
Steve Silverman
Susannah Weaver
Consultants for Environmental Defense Fund

[3] *See* discussion *infra* Section VII.

TABLE OF CONTENTS

OVERVIEW

The Proposal acknowledges that "[t]he best available science must serve as the foundation of EPA's regulatory actions."[4] But it then requires EPA to systematically ignore the best available science when it regulates to protect human health and welfare. This is counter to EPA's statutory mandates to use "best available science," and the proposal is a transparent attempt not to *strengthen* science, but rather to *censor* science that is inconvenient to the current Administration's political goals.

Since EPA was established nearly half a century ago, the Agency and its leadership—under Administrations of both parties—have recognized the central role that rigorous science plays in fulfilling the Agency's mission of protecting human health and the environment.[5] EPA's obligation to consider the best available science is not only a policy commitment that flows from the Agency's mission; it is a legal obligation enshrined in many of the fundamental public health and environmental statutes that EPA is charged with administering. The agency has established an array of mechanisms over the last five decades—including "rigorous review" by its scientific advisory boards "that goes beyond the typical journal peer review procedures"[6]—to ensure that the Agency's decisions are grounded in the best available science.

The Administrator's proposal does not build on this strong foundation; to the contrary, it crumbles it. The purpose and effect of the proposal would be to *degrade* the quality of science in EPA's decision making. While the proposal suggests that its aim is to improve transparency by increasing public availability of data, in actuality it proposes none of the steps that a proposal seriously aimed at that goal would propose, such as increasing funding for EPA grantees to undertake this effort, or proposing solutions to real concerns about patient confidentiality. Instead, the heart of the proposal is a bar on considering science simply because the underlying data is not publicly available, regardless of whether the science has been peer reviewed, reproduced, or contains other hallmarks of scientific quality. Indeed, the agency's recent communication to the Congressional Budget Office that a similar Congressional proposal could be implemented at "no cost" proves the point: EPA's aim here is not to make more data available (which costs money), but to rely on less science in decisionmaking.

The agency's arbitrary, single-minded focus on considering studies for which certain data and models are publicly available (but only the dose-response studies relevant to health

---

[4] 83 Fed. Reg. at 18,769.

[5] Brady Dennis, *Outgoing EPA chief: Science is 'fundamental to absolutely everything we do'*, Washington Post (Dec. 21, 2016) (quoting former EPA Administrator Gina McCarthy as saying, "Science is everything. Almost every action we take is bounded by what the science tells us. It's based on a factual record of where the world is today and what is our obligation under our mission. Science needs to be protected. Any effort to undermine that science in a way that would give undue influence to folks that aren't scientists is a really big problem."), https://www.washingtonpost.com/news/energy-environment/wp/2016/12/21/outgoing-epa-chief-science-is-everything-it-is-fundamental-to-absolutely-everything-we-do/?utm_term=.6f1e45472169; Christine Todd Whitman, *No room for science in Trump Administration,* CNN (May 15, 2017), https://www.cnn.com/2017/05/15/opinions/no-science-in-trump-administration-whitman/index.html (describing Administrator Pruitt's actions as a "trend away from science as the backbone of the EPA and other key federal agencies").

[6] Memorandum by Alison Cullen, Chair, SAB Work Group on EPA Planned Actions for SAB Consideration of the Underlying Science 4 (May 12, 2018) (observing that the Proposal "fails to mention that EPA has mechanisms for vetting science through several expert panels," including the SAB and others).

protective regulation, not the ones supporting registration of chemicals) stands in stark contrast to the way the scientific community validates research findings. The scientific community, and scientific journals look to a range of attributes when assessing the quality of a scientific study, including whether the study has been peer reviewed, whether the scientists used rigorous scientific methods, and whether the study's results have been reproduced or replicated. While scientific journals and other institutions have encouraged making data and models publicly available, there is widespread recognition in the scientific community that doing so is often legitimately constrained due to legal and ethical protections on the confidentiality and privacy of data, or because the data is unavailable. Moreover, no scientist or scientific organization supports the Proposal's approach of *excluding* research for which the underlying data cannot be disclosed. Indeed, *none* of the materials EPA cites support such an extreme approach. To the contrary, the scientific community recognizes that the quality of a study is not determined by whether the underlying data is publicly available and has long utilized a variety of tools for ensuring the integrity and rigor of research findings.[7]

For all these reasons, numerous representatives of the scientific community—including editors of the very scientific journals whose policies EPA cites to in the Proposal, the American Association for the Advancement of Science, members of the SAB, and other scientists cited to by EPA—have already voiced serious concerns about the Proposal.[8] As these experts have recognized, it is not consistent with good scientific practice, and certainly not consistent with the Agency's responsibility to utilize "best available science," to deem certain scientific studies unworthy of consideration simply because these studies cannot meet an arbitrary public availability requirement.[9] Far from promoting the integrity of Agency decisions, the Proposal's simplistic approach would impoverish the Agency's decision-making by excluding the consideration of scientific studies that, standing alone or in combination with other studies, have significant bearing on vital public health and environmental protections. This, in turn, would result in regulations that are *not* based on "best available science" and that will provide inadequate protection for the very public health and welfare that EPA has been charged by Congress to safeguard.

---

[7] *See id.* at 4 ("The proposed rule fails to mention that there are various ways to assess the validity of prior epidemiologic studies without public access to data and analytic methods.").

[8] *E.g.*, Anne Q. Hoy, *Scientific Leaders Speak Out on EPA's Proposed "Transparency Rule,"* https://www.aaas.org/news/scientific-leaders-speak-out-epa-s-proposed-transparency-rule; Jeremy Berg et al., *Joint Statement on EPA Proposed Rule and Public Availability of Data,* Science (Apr. 30, 2018), http://science.sciencemag.org/content/early/2018/04/30/science.aau0116; Letter to Acting Administrator Wheeler from Marcia McNutt, President of the National Academy of Sciences, C.D. Mote, Jr., President of the National Academy of Engineering, and Victor J. Dzau, President of the National Academy of Medicine (July 16, 2018) (Warning that "overly stringent requirements for transparency may cause valid evidence to be discarded and thereby pose a threat to the credibility of regulatory science," and stating that "The National Academies have developed a long-standing body of work that demonstrates scientific literature can be evaluated in a transparent and objective manner without complete disclosure of the underlying data.").

[9] *See* John Ioannidis, *All science should inform policy and regulation*, 15 PLOS 5 (May 3, 2018) ("Past collected and analyzed information can and should still be used for decision-making, taking into account any relevant imperfections. While fully transparent and reproducible information should certainly be valued more highly, studies with weaknesses can still offer insights.").

That, of course, appears to be the current Administration's goal. A close examination of the history of this Proposal confirms that its purpose is not to strengthen science at EPA, but to undermine public health and environmental protections by arbitrarily blinding the agency to vital research. Indeed, the Proposal resembles proposals advanced by the tobacco industry for the specific purpose of suppressing public health science warning about the dangers of tobacco smoke.[10] The Proposal also resembles failed legislation in Congress that was similarly advanced by industry interests seeking to undermine public health and environmental protections, and criticized by scientific experts.[11] EPA documents released in response to Freedom of Information Act (FOIA) requests relating to the Proposal show that Trump Administration appointees deliberately tailored the scope of the Proposal in order to promote industry interests.

EPA's purpose and mission is to protect human health and welfare, *not* to promote the agendas of the worst polluters and their allies in order to weaken health and welfare protections. EPA should withdraw this misguided and harmful proposal.

Terminology

At the outset, it is useful to review relevant terminology, which the Proposal appears to confuse and conflate. A recent National Academy of Sciences workshop produced the following definitions of "reanalysis," "replication," and "reproduction," each of which has a different scientific meaning and different applications and implications.[12] Let's consider each of these definitions separately.

> *A **reanalysis** is when you conduct a further analysis of data. A person doing a reanalysis of data may use the same programs and statistical methodologies that were originally used to analyze the data or may use alternative methodologies, but the point is to analyze exactly the same data to see if the same result emerges from the analysis.*

A reanalysis does validate or invalidate a study findings. If all credible methods of reanalysis yield effectively the same results as the original analysis, this does strengthen the original findings. The use of differing statistical models should be assessed with care and demonstrate that the assumptions supporting a new method of analysis is significantly more credible than the original analysis. It is easy to develop methods of analysis that can demonstrate

---

[10] Emily Atkin, *The EPA is Acting Like Big Tobacco,* The New Republic (Apr. 26, 2018), https://newrepublic.com/article/148126/epa-acting-like-big-tobacco (describing the role of Steve Milloy, a leading public proponent of the Proposal who has taken credit for its existence, in crafting similar policy proposals on behalf of the tobacco industry-funded Advancement of Sound Science Coalition).

[11] Letter by U.S. Science, Engineering, and Academic Institutions to Kevin McCarthy, House Majority Whip (Mar. 16, 2015) (opposing "Secret Science Reform Act, H.R. 1030"), https://sciencepolicy.agu.org/files/2013/07/AAAS-Secret-Science-letter-McCarthy-2015.pdf; Letter by Barry Nussbaum, American Statistical Association to Sen. Mike Rounds and Sen. Kamala Harris (May 25, 2017) (opposing HONEST Act, H.R. 1430), https://www.amstat.org/asa/files/pdfs/POL-HONEST_ActLetter.pdf.

[12] National Academies of Sciences, Engineering, and Medicine, *Principles and obstacles for sharing data from environmental health research: Workshop summary*, The National Academies Press (2016), https://www.nap.edu/catalog/21703/principles-and-obstacles-for-sharing-data-from-environmental-health-research.

a different finding, but are created solely for that purpose and these should not be given greater weight in evaluating a particular study.

> ***Replication*** *means that you actually repeat a scientific experiment or a trial to obtain a consistent result. The second experiment uses exactly the same protocols and statistical programs but with different data from a different population*[13]*. The goal is to see if the same results hold with data from a different population.*

Replication predominantly applies to laboratory studies and randomized control trials since you are able to control almost all of the experimental details making replication possible. Replication does not enhance transparency. In environmental epidemiology, randomized control trials are not feasible or ethical, and replication of observational studies is virtually impossible since it is not possible to create the same conditions as seen in the original study. Even in laboratory experiments, replication can be difficult due to uncontrolled factors like genetic drift in cell lines and animal strains. Finally, if you do have replicate studies and one has a positive finding and another has a negative finding, there would have to be additional criteria used to determine which study was correct; thus a failure to replicate should not immediately lead to the conclusion that there is no effect. Rather than replicating a study, it is far better to develop a better study that replicates the results while providing greater insight into the basis underlying any toxicity.

> *And then, finally, when you* ***reproduce*** *a scientific experiment, you are producing something that is very similar to that research, but it is in a different medium or context. For example, a researcher who is reproducing an experiment addresses the same research question but from a different angle than the original researcher did.*

Here, reproduction refers to a body of evidence addressing the same hypothesis, but using different populations, methods, etc. Reproduction does not enhance transparency. The majority of research on the health effects of environmental hazards fall into this category. Here, a series of studies that address the same hypothesis and give the same basic result does indeed strengthen findings of toxicity.

None of these concepts discusses the scientific quality of the study; this is critical. The ability to replicate a study with very poor scientific quality does not strengthen the scientific belief that any toxicity is present. Similarly, studies that attempt to reproduce the same findings must have their quality clearly established before comparisons can be made across the multiple studies.

An example of how some of these different techniques work in practice is the scientific evidence on air pollution and premature death which include the Harvard Six Cities Study and the American Cancer Society Cancer Prevention Study II (ACS CPSII). The extent to which these studies have been reanalyzed and reproduced is extraordinary and by no means necessary. But they provide a good case study of how these techniques work in practice.

---

[13] "Different population" in this context means a different set of the same test subjects (e.g., same animal species and strain, same cell lines).

The original Harvard Six Cities and ACS CPSII studies on mortality were published in 1993 and 1995 respectively.

- The Harvard Six Cities study assessed the long-term effects of fine particle pollution (PM2.5) over 12 to 14 years (1974–1989) on premature mortality among 8,111 adult participants who lived in 6 different cities: Watertown, MA; Harriman, TN; St. Louis, MO; Steubenville, OH; Portage, WI; and Topeka, KS. After accounting for cigarette smoking, level of education, body mass index, and occupational exposure to dusts, gases, and fumes, the authors of this study found that for members of the same age and sex group there was a 26% higher risk of premature mortality between the study participants living in the city with the highest levels of particles (Steubenville) and the city with the lowest levels (Portage).[14]

- The investigators of the Harvard Six Cities study, along with others, **reproduced** their finding in a separate assessment of the association between fine particle levels and mortality among 295,223 adults who lived in 50 metropolitan areas across the United States, over a period of 7 years (1979-1983) in the ACS CPSII study. After accounting for smoking, education, body mass index, alcohol consumption, and self-reported occupational exposure to a number of substances, the scientists found that for participants of the same age, race and sex there was a 17% increased risk of mortality with every 25.4 microgram per meter cube change in PM2.5.[15]

The Harvard Six Cities Study and the ACSCPSII were **reanalyzed** by the Health Effects Institute, a nonprofit independent research corporation funded by EPA and the motor vehicle industry, under a data sharing agreement. A research team evaluated the consistency and accuracy of the data and then undertook a series of comprehensive analyses to test the validity of the findings first using the same statistical analyses and then testing the robustness of the original findings and interpretations to alternative analytic approaches. The results of the reanalysis were resoundingly similar to the original studies. For the Harvard Six cities study the reanalysis found a 28% increased risk of mortality per 18.6 microgram per meter cube of PM2.5 in comparison to 26% found in the original study. For the ACS CPSII study the showed that for every 25.4 microgram per meter cube change in PM2.5 there was an associated 18% increased risk of mortality (results of the independent reanalysis) vs 17% reported by the original study.[16]

---

[14] Dockery, D.W., Pope, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris Jr, B.G. and Speizer, F.E., *An Association Between Air Pollution and Mortality in Six US Cities*, 329(24) New England Journal of Medicine 1753-1759 (1993).

[15] Pope, C.A., Thun, M.J., Namboodiri, M.M., Dockery, D.W., Evans, J.S., Speizer, F.E. and Heath, C.W., *Particulate Air Pollution as a Predictor of Mortality in a Prospective Study of US Adults*, 151(3) American Journal of Respiratory and Critical Care Medicine 669-674 (1995).

[16] Krewski, Daniel, et al., *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of particulate air pollution and mortality*, footnote on 249 Health Effects Institute (2000). See also Letter to Andrew Wheeler from Harvard University (Docket ID No. EPA-HQ-OA-2018-0259) (reanalysis and "releasing raw data will not improve the quality of the resulting report/study/analysis, and therefore will do nothing to render any individual study 'better.'").

A large body of literature also shows that this association of fine particle pollution and mortality has been **reproduced** in different populations across the globe,[17] over different periods of time, contexts and using different methods. Most recently, a study of 61 million elderly people enrolled in Medicare across the entire United States followed over 13 years found a strong association between particle pollution and increased risk of mortality, at even the current levels of air pollution and below the current air quality standards for PM2.5.[18] It is this accumulation of evidence of reproducible effects in multiple studies that is critical in determination of causality and validation of an effect and is already an integral part of the EPA process of supporting causality.[19]

Through these different methods, the original findings of the Harvard Six Cities Study have been validated many times over, and they have been used to inform countless EPA rule makings that address particulate matter pollution. Notably, however, the Proposal would appear to preclude EPA from using them because—while the Study has been reanalyzed and reproduced—the underlying data is not publicly available because of patient confidentiality protections bound by individual contractual agreements between the scientists and the research participants and by the Health Insurance Portability and Accountability Act. These reasons are unrelated to the validity, integrity or quality of the Harvard Six Cities Study. Indeed, the Office of Management and Budget's data quality guidelines specifically point to the Harvard Six Cities Study as an example of how data may be validated or corroborated without public release of the underlying raw data.[20] It is critically important to note that reanalysis projects are not simple or inexpensive.[21] The reanalysis of just the Harvard Six Cities Study and the ACS CPSII took three years to complete and cost $899,046 in direct expenditures,[22] without accounting for costs incurred by Health Effects Institute for oversight and review as well as staff compensation.

In summary, reanalysis is a tool to demonstrate the robustness of an effect to changes in the statistical model underlying an analysis of a single data set. However, it is easy to develop methods of reanalysis that can demonstrate a different finding. Therefore, care must be taken to understand the assumptions underlying models applied in reanalysis in order to judge their relevance. Replication in the environmental health context is primarily limited to laboratory studies and, without additional information to guide a decision, provides little information that can be used to decide between replicate studies with differing results. Reproducing effects in multiple studies that are not identical is the basis for almost all scientific decisions on environmental issues and should be the focus of the EPA's approach to regulatory science. Finally, none of these issues address other key aspects of scientific quality such as

---

[17] EPA, NCEA, *Integrated Science Assessment for Particulate Matter*, EPA/600/R-08/139F (2009); Beelen, Rob, et al., *Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project*, 383.9919 The Lancet 785-795 (2014).

[18] Di, Qian, et al., *Air pollution and mortality in the Medicare population*, 376.26 New England Journal of Medicine 2513-2522 (2017).

[19] EPA, *Preamble to the Integrated Science Assessments (ISA)* (EPA/600/R-15/067) (2015).

[20] *OMB's Guidelines Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information*, 67 Fed. Reg. 8,452, 8,456 (Feb. 22, 2002).

[21] Comments of Daniel Greenbaum, President, Health Effects Institute (HEI), on Proposed Rule EPA–HQ–OA–2018–0259 (July 17, 2018).

[22] Krewski, Daniel, et al., *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of particulate air pollution and mortality*, footnote on 249 Health Effects Institute (2000).

generalizability and bias; how these characteristics of any scientific study are assessed by the EPA directly relate to the transparency of any decisions they might make.

I. EPA's Proposed Rule Violates Numerous Substantive Statutory Requirements.

A. EPA Does Not Have Authority to Issue the Proposed Rule.

Agencies are creatures of Congress; "an agency literally has no power to act . . . unless and until Congress confers power upon it." *Louisiana Pub. Serv. Comm'n v. FCC*, 476 U.S. 355, 374 (1986); *see Am. Library Ass'n v. FCC*, 406 F.3d 689, 691 (D.C. Cir. 2005) ("It is axiomatic that administrative agencies may issue regulations only pursuant to authority delegated to them by Congress."). EPA points to a smattering of statutes as allegedly authorizing the Proposal.[23] None of these authorities, however, authorize EPA to promulgate a one-size-fits-all regulation governing how the agency will consider science under its various statutory authorities, which is perhaps why EPA solicits comment on whether additional authorities might exist to authorize its Proposal. The varied statutes that the Proposal cites have different requirements as to the agency's obligations when considering science. *Compare* CAA § 108(a) (standards must "reflect the latest scientific knowledge useful in indicating" health and welfare effects)[24] *with* TSCA § 4(f) (Administrator must consider "*any other information available*")[25] *with* Safe Drinking Water Act ("SDWA") § 1412(b)(1)(B)(ii)(II) (Administrator must consider "the best available public health information").[26] The Proposal gives no explanation of how *any* of the provisions it cites provide authority for the Proposal, much less how all of them authorize identical requirements.

For example, EPA cites the Clean Air Act, § 301, 42 U.S.C. § 7601, as purportedly granting authority for the Proposal.[27] The authority granted by section 301(a), however, applies only to the Clean Air Act and, in any event, is not broad enough to encompass this Proposal. Section 301 provides that "[t]he Administrator is authorized to prescribe such regulations subject to section 307(d) as are *necessary* to carry out his [or her] functions under this Act."[28] The courts have consistently "decline[d] to read … open-ended power into section 301,"[29] and instead have required that regulations promulgated under section 301 be both necessary and appropriate.[30] As

---

[23] 83 Fed. Reg. at 18769.
[24] 42 U.S.C. § 7408(a).
[25] 15 U.S.C. § 2603(f).
[26] 42 U.S.C. § 300g-1(b)(1)(B)(ii)(II), (b)(1)(A)(i); *see also*, 42 U.S.C. § 300g-1(b)(3)(A)(i) ("the Administrator shall use. . . the best available, peer-reviewed science and supporting studies conducted in accordance with sound and objective scientific practices").
[27] 83 Fed. Reg. at 18769.
[28] 42 U.S.C. § 7601(a)(1) (emphasis added).
[29] *Nat. Res. Def. Council v. Reilly*, 976 F.2d 36, 41 (D.C. Cir. 1992).
[30] *E.g.*, *Alabama Power Co. v. Costle*, 636 F.2d 323, 403 (D.C. Cir. 1979) (finding an EPA rule unauthorized under section 301, and concluding that "[a]n extension of PSD permit requirements beyond the wording of the Act is therefore neither necessary nor appropriate to carry out EPA's functions under the Act."); *Nat. Res. Def. Council v. EPA*, 22 F.3d 1125, 1148 (D.C. Cir. 1994) ("[S]ection 301 does not provide the Administrator '*carte blanche* authority to promulgate any rules, on any matter relating to the Clean Air Act, in any manner that the Administrator wishes,'" and instead "allow[s] the promulgation of rules that are necessary and reasonable to effect the purposes of

discussed in more detail below, EPA's Proposal here is not necessary, and instead directly conflicts with several other provisions of the Clean Air Act. It is axiomatic that a "general grant of authority cannot trump specific statutory provisions."[31]

Nor does Congressional authorization to *conduct* or *fund* research authorize EPA to *ignore* research in regulatory decision-making. Accordingly, provisions like TSCA § 10, which directs that the "Administrator shall … conduct such research, development, and monitoring as is necessary to carry out the purposes of this [Act],"[32] and CAA § 103, which authorizes the agency to conduct and support research,[33] plainly do not authorize the Proposal.

B. The Proposed Rule Violates EPA's Statutory Authorities.

Not only is there no authority for EPA's pan-statutory Proposal, the Proposal would violate explicit statutory commands. Though EPA admits that "[t]he best available science must serve as the foundation of EPA's regulatory actions,"[34] proposed section 30.5 would ***prohibit*** EPA from considering high quality and critically important scientific studies—precisely that "best available science"—when undertaking regulatory actions. Specifically, section 30.5 would prevent EPA from considering any scientific study for which the underlying "dose response data and models" are not "publicly available in a manner sufficient for independent validation."[35] This would be true even if that scientific study constituted "information available to the Administrator" in a TSCA § 4(f) rulemaking, 15 U.S.C. § 2603(f)(2); "reflect[ed] the latest scientific knowledge useful in indicating" health and welfare effects in a CAA § 108 rulemaking, 42 U.S.C. § 7408(a)(2); or reflected "the best available public health information" in a SDWA rulemaking, 42 U.S.C. § 300g-1(b)(1)(B)(ii)(II). Accordingly, this proposed prohibition would contravene an array of statutes governing EPA's consideration of science when promulgating rules, such as requirements to consider the "best available science" when setting environmental protection standards. *See, e.g.,* SDWA, 42 U.S.C. § 300g-1(b)(3)(A) (EPA must use "[t]he best available, peer-reviewed science and supporting studies conducted in accordance with sound and objective scientific practices" and "[d]ata collected by accepted methods or best available methods"); TSCA, 15 U.S.C. § 2625(h) ("[T]he Administrator shall use scientific information, technical procedures, measures, methods, protocols, methodologies, or models, employed in a manner consistent with the best available science."); CAA, 42 U.S.C. § 7408(a) (EPA shall establish air quality criteria that "shall accurately reflect the latest scientific knowledge useful in indicating the kind and extent of all identifiable effects on public health or welfare which may be

---

the Act.") (quoting *Citizens to Save Spencer County v. EPA*, 600 F.2d 844, 873 (D.C. Cir. 1979)); *Nat. Res. Def. Council v. EPA*, 749 F.3d 1055, 1063 (D.C. Cir. 2014) ("[W]e have consistently held that EPA's authority to issue ancillary regulations is not open-ended, particularly when there is statutory language on point."); *North Carolina v. EPA*, 531 F.3d 896, 922 (D.C. Cir. 2008), *on reh'g in part*, 550 F.3d 1176 (D.C. Cir. 2008) (striking down a regulation promulgated under Section 301 because EPA could not demonstrate that it was "necessary" to fulfill the purposes of the Act).

[31] *Nat. Res. Def. Council v. EPA*, 749 F.3d 1055, 1063-64 (D.C. Cir. 2014); *API v. EPA*, 52 F.3d 1113, 1119 (D.C. Cir. 1995) (same).

[32] 15 U.S.C. § 2609(a), cited at 83 Fed. Reg. at 18769.

[33] 42 U.S.C. § 7403, cited at 83 Fed. Reg. at 18769.

[34] 83 Fed. Reg. at 18769.

[35] 83 Fed. Reg. at 18773-74.

expected from the presence of such pollutant in the ambient air, in varying quantities."). And, by excluding science that meets these statutory criteria from supporting regulations to protect public health and welfare, the Proposal would frustrate Congress's policy in these statutes and frustrate EPA from achieving its fundamental mission.[36]

1. EPA's statutory authorities generally require the agency to consider *all* available data when undertaking significant rulemakings.

As just noted, EPA's statutory authorities mandate a variety of requirements for what scientific information EPA must consider in rulemaking. These statutes are discussed in detail, *infra* at Section I.B.3. To take one example that appears in numerous statutes, including TSCA, CAA, SDWA, and the Endangered Species Act, Congress has often required agencies to act on the "best available science." For an agency to comply with this obligation, the agency must at least consider all available scientific information. "Best" means "of the most excellent, effective, or desirable type or quality."[37] "Available" means "able to be used or obtained."[38] And "science" means "the intellectual and practical activity encompassing the systematic study of the structure and behavior of the physical and natural world through observation and experiment."[39] Assessing which science is "best" requires consideration of the overall quality of the science, and the public availability of underlying data is, at best, one of many aspects that should inform that assessment of overall quality.

An agency "cannot ignore available. . . information."[40] Numerous courts have indicated that a plaintiff or petitioner can establish a violation of the "best available science" requirement by "point[ing] to any scientific evidence that the agency failed to consider."[41] "The best available data requirement. . . prohibits [an agency] from disregarding available scientific evidence that is in some way better than the evidence [it] relies on.'"[42] "An agency does. . . have an obligation to deal with newly acquired evidence in some reasonable fashion."[43] EPA's proposal will result in EPA precluding itself from considering certain studies that are "available," thus violating the requirement that EPA rely on the best available science.

In addition, the requirement that agencies use "best available" science or information often means that the agency must act *even if* the available science or information is imperfect.

---

[36] *See, e.g., Shays v. FEC*, 528 F.3d 914, 919 (D.C. Cir. 2008) ("[W]e 'must reject administrative constructions of [a] statute that frustrate the policy that Congress sought to implement.'") (*quoting Cont'l Air Lines, Inc. v. Dep't of Transp*, 843 F.2d 1444, 1453 (D.C. Cir. 1988)).

[37] *Oxford American Dictionary* 159 (3d ed. 2010).

[38] *Id.* at 111.

[39] *Id.* at 1564.

[40] *Conner v. Burford*, 848 F.2d 1441, 1454 (9th Cir. 1988); *San Luis & Delta-Mendota Water Auth. v. Jewell*, 747 F.3d 581, 602 (9th Cir. 2014) (quoting *Kern Cnty.*, 450 F.3d at 1080-81 (quoting *Conner v. Burford,* 848 F.2d 1441, 1454 (9th Cir. 1988)).

[41] *Safari Club Int'l v. Salazar (In re Polar Bear Endangered Species Act Listing & Section 4(d) Rule Litig. - MDL No. 1993)*, 709 F.3d 1, 9 (D.C. Cir. 2013).

[42] *Kern Cty. Farm Bureau v. Allen*, 450 F.3d 1072, 1080 (9th Cir. 2006) (quoting *Sw. Ctr. for Biological Diversity v. Babbitt*, 215 F.3d 58, 60 (D.C. Cir. 2000)).

[43] *Catawba County v. EPA*, 571 F.3d 20, 45 (D.C. Cir. 2009) (quoting *American Iron & Steel Institute v. EPA*, 115 F.3d 979, 1007 (D.C. Cir. 1991)).

"Even if the available scientific and commercial data were quite inconclusive, [the agency] may—indeed must—still rely on it" when the agency has a duty to act.[44] "[W]here the information is not readily available, we cannot insist on perfection."[45] Just as the Courts have recognized that they cannot expect perfection, agencies cannot choose to ignore certain studies or sources of information based solely on whether the data is publicly available—especially where the validity of those studies has been established using techniques that do not rely on public availability of underlying data.

EPA cannot reasonably elevate the interest in public availability of all underlying information above all other factors in assessing the "best available science." Textually, EPA's approach is unlawful.

2. The proposal violates these statutory commands by requiring EPA to ignore science when undertaking significant rulemakings.

In direct violation of statutory requirements to consider, for example, "any other information available" or "the latest scientific knowledge [that is] useful" or "best available science," the Proposal would *prohibit* EPA from considering relevant and high quality science whenever the underlying data for a study is not publicly available. Through the Proposal, EPA unlawfully tries to engraft an additional statutory requirement onto each of these statutes, requiring that to be considered a study's underlying data must be publicly available.[46] For EPA's Proposal to succeed, EPA must demonstrate that a study *cannot* be "other information available to the Administrator" or the "latest scientific knowledge useful in indicating" health or welfare effects or the "best available science," or any of a number of other statutory formulations if the underlying data is not publicly available. EPA's Proposal fails to do so, and it could not do so.

As explained *infra* at Section II.A.1, there are many reasons that underlying study data may not be available that have no bearing on the quality or validity of the study. These include legal restrictions or concerns about privacy (especially with respect to studies involving human subjects), confidentiality, confidential business information, or national security. Further, if this requirement were applied retroactively to existing studies, it may no longer be possible to make underlying data and models publicly available. EPA acknowledges these impediments in proposed section 30.9, which provides the Administrator with discretion—but not an obligation—to allow the agency to consider a study for which underlying data or models are not publicly available if he determines that public disclosure is infeasible. But where the Administrator fails to exercise his discretion to grant an exemption pursuant to proposed section 30.9, or where data or models are unavailable for reasons that do not satisfy the infeasibility standard, proposed section 30.5 would prohibit EPA from considering such studies, regardless of whether they meet the statutory criteria for consideration.

The only way that this prohibition could comport with EPA's statutory obligations is if a study for which underlying data is not available *cannot* be, for example, "other information

[44] *Southwest Ctr. for Biological Diversity v. Babbitt*, 215 F.3d 58, 60 (D.C. Cir. 2000) (quoting *City of Las Vegas v. Lujan*, 891 F.2d 927, 933 (D.C. Cir. 1989)).
[45] *San Luis*, 747 F.3d at 602.
[46] *See Nat'l Ass'n of Homebuilders v. Defenders of Wildlife*, 551 U.S. 644, 663-64 (2007).

available" or "the latest scientific knowledge [that is] useful" or "best available science"—i.e., if the public unavailability of a study's underlying dose response data and models makes the study ineligible to meet these criteria, regardless of whether the study has been peer reviewed, is based on rigorous methodologies, or has been published in a leading journal, and regardless of the reason for the public unavailability. EPA makes no such demonstration—nor could it. There is simply no support for such a proposition; to the contrary, all of the evidence shows that studies may be "best available science," and certainly "other information available" regardless of whether the data underlying them is publicly available.

What the Proposal fails to recognize is that disclosure of data addresses only *one* method of validating scientific research—and a relatively less important aspect at that. Disclosure of data for a given study—the focus of the Proposal—permits independent researchers to determine whether the data and methodology *used in that study* can be applied to generate the *same* results. This may help protect against sources of error or misrepresentation in a particular study. However, both EPA and independent researchers have recognized that such reanalysis does not by itself *validate* a particular study.[47] Rather, a study's evidentiary weight rests both on the strength of its methodology, as well as whether similar results can be obtained by applying the study's methodology to a relevant, but *different* dataset or population, or by using a distinct methodology to interrogate the same hypothesis.[48]

a) The scientific community

Publication in a peer-reviewed scientific journal is the way that scientists communicate their findings to other scientists and is considered the hallmark of scientific quality. Notably, the editors in chief of the world's top scientific journals have notified EPA that "[i]t does not strengthen policies based on scientific evidence to limit the scientific evidence that can inform them; rather, it is paramount that the full suite of relevant science vetted through peer review, which includes ever more rigorous features, inform the landscape of decision making."[49] In response to EPA's Proposal, the editors-in-chief of *Science* and *Nature*, and other leading scientists explained that though "[d]ata sharing is a feature that contributes to the robustness of published scientific results. . . in not every case can all data be fully shared."[50] For example, full

---

[47] *See* EPA, *Preamble to the Integrated Science Assessment* at 20 (2015) ("An inference of causality is strengthened when a pattern of elevated risks is observed across several independent studies. *The reproducibility of findings constitutes one of the strongest arguments for causality*. . . ") (emphasis added); National Academies, *Principles and Obstacles for Sharing Data From Environmental Health Research* 6 (2016) (quoting researcher Lynn Goldman's observation that reproducibility and replicability across independent studies – as distinct from reanalysis of a single set of data using the same methodology – are the most convincing ways of validating a research finding); Lynn R. Goldman & Ellen Silbergeld, *Correspondence on Access to Chemical Data Used in Regulatory Decision Making*, 121 Environmental Health Perspectives A111 (Apr. 2013), https://ehp.niehs.nih.gov/wp-content/uploads/121/4/ehp.1206438.pdf ("Replication in science is quite different; it involves performance of an independent study with the same hypothesis and then testing the extent to which this independent study reaches the same conclusions. . . Designing and conducting a replication study does not require access to raw data from the original study; this would abrogate the concept of independence.")

[48] *See* National Academies, *Principles and Obstacles* at 6.

[49] Jeremy Berg et al., *Joint Statement on EPA Proposed Rule and Public Availability of Data,* Science (Apr. 30, 2018), http://science.sciencemag.org/content/early/2018/04/30/science.aau0116.

[50] *Id.*

sharing is not possible when data sets include "personal identifiers."[51] The scientists confirm that even under circumstances where underlying data cannot be made generally available, it is possible to evaluate the merits of a study, explaining:

> Importantly, the merits of studies relying on data that cannot be made publicly available can still be judged. Reviewers can have confidential access to key data and as a core skill, scientists are trained in assessing research publications by judging the articulation and logic of the research design, the clarity of the description of the methods used for data collection and analysis, and appropriate citation of previous results.[52]

They conclude that EPA's proposal to exclude relevant studies from EPA's consideration based solely on the fact that underlying data or methods cannot be made available to the public "will adversely affect decision-making processes."[53]

In a letter filed in this docket, the Presidents of the National Academies of Science, Engineering, and Medicine similarly observe that the public availability of data is not necessary to ensure the integrity of regulatory science and is not a sufficient criterion for excluding a particular study from consideration. The Presidents' letter notes: "The National Academies have developed a long-standing body of work that demonstrates scientific literature can be evaluated in a transparent and objective manner without complete disclosure of the underlying data." [54] The letter goes on to explain: "If the study data are not available, their absence may affect how the study is rated and used in the [agency's] analysis, but the study should not necessarily be eliminated from the assessment."[55]

b) EPA policy and practice

EPA has previously stated in several different forums that a scientific study can be valid even if the underlying dose response data and models are not publicly available. For example, EPA recently explained in its own *Plan to Increase Access to Results of EPA-Funded Scientific Research* that even though "some research data cannot be made fully available to the public but instead may need to be made available in more limited ways," the lack of full public availability "does not affect the validity of the scientific conclusions from peer-reviewed research publications."[56] Under the plan, EPA must make publications resulting from EPA-funded research publicly accessible on National Institute of Health's PubMed Central (PMC).[57] The plan

---

[51] *Id.*
[52] *Id.*
[53] *Id.*
[54] Letter to Acting Administrator Wheeler from Marcia McNutt, President of the National Academy of Sciences, C.D. Mote, Jr., President of the National Academy of Engineering, and Victor J. Dzau, President of the National Academy of Medicine 2 (July 16, 2018), http://www.nationalacademies.org/includes/EPA%20Proposed%20Rule%20Docket%20EPA-HQ-OA-2018-0259%20NASEM%20Comment.pdf.
[55] *Id.* at 2-3.
[56] EPA, *Plan to Increase Access to Results of EPA-Funded Scientific Research* 4-5 (Nov. 29, 2016), https://www.epa.gov/sites/production/files/2016-12/documents/epascientificresearchtransperancyplan.pdf.
[57] *Id.* at 8.

aims to "maximize access, by the general public and without charge, to digitally formatted data resulting from EPA funded research, *while protecting confidentiality and personal privacy, recognizing proprietary interests, business confidential information and intellectual property rights, and preserving the balance between the relative benefits and costs of long-term preservation and access.*"[58] The plan recognizes important exceptions for when "the research data cannot be released due to one or more constraints, such as requirements to protect confidentiality, personal privacy, proprietary interest, or property rights."[59] It specifically declares: "The validity of scientific conclusions drawn from research publications or their associated research data, or EPA's ability to consider those conclusions and data in its actions, does not depend on compliance with this Plan."[60]

Likewise, EPA's Science Policy Council explains in *A Summary of General Assessment Factors for Evaluating the Quality of Scientific and Technical Information* that EPA's determination as to the quality and reliability of a particular scientific study does not depend on one single factor (e.g., the public availability of underlying data), but instead turns on the agency's consideration of five general factors.[61] Congress implicitly endorsed this approach by including a directive for EPA to use these same five factors in evaluating science under the Toxic Substances Control Act Amendments passed in 2016,[62] and just last year this Administration included these same factors in a recent regulation implementing TSCA.[63] The factors comprise: (1) soundness; (2) applicability and utility; (3) clarity and completeness; (4) uncertainty and variability; and (5) evaluation and review.[64] Of these, the only ones with any possible direct relevance to EPA's proposed approach are the third and fifth factors, but neither supports the elevation of public availability of data above all other considerations or the exclusion of studies with non-public data. The third factor, "clarity and completeness" requires EPA to consider "[t]he degree of clarity and completeness with which the data, assumptions, methods, quality assurance, sponsoring organizations and analyses employed to generate the information are documented." The fifth factor, "evaluation and review," requires EPA to consider "[t]he extent of independent verification, validation and peer review of the information or of the procedures, measures, methods or models." Even clear and complete "documentation" of the data used does not require that the data be made publicly available. Nor does factor five require either that a study's findings must have been replicated using the same data, or that the data must be available

---

[58] *Id.* at 11 (emphasis added).
[59] *Id.*
[60] *Id.* at 6.
[61] EPA Science Policy Council, *A Summary of General Assessment Factors for Evaluating the Quality of Scientific and Technical Information*, EPA 100/B-03/001 (June 2003) https://www.epa.gov/risk/summary-general-assessment-factors-evaluating-quality-scientific-and-technical-information.
[62] *Id.* at 7.
[63] EPA Science Policy Council, *A Summary of General Assessment Factors for Evaluating the Quality of Scientific and Technical Information*; 15 U.S.C. § 2625(h)(1)-(5); 82 Fed. Reg. 33,726, 33,731 (July 20, 2017), 42 U.S.C. § 300g-1(b)(3)(A).
[64] Note that TSCA and the regulations do not include the headers for the five factors ("soundness," "applicability and utility," etc.) included in the Science Policy Council guidance, but the description of each factor to be considered is largely identical.

to allow for such replication. Moreover, these are only portions of two of five key factors to consider.[65]

Similarly, EPA's *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of the Information Disseminated by the Environmental Protection Agency*,[66] ("EPA Information Quality Guidelines") issued pursuant to Section 515(a) of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106-554; H.R. 5658) (the "Data Quality Act") make it clear that the public unavailability of underlying data or models does not render a study inappropriate for EPA's consideration. Specifically, the EPA Information Quality Guidelines acknowledge that even with respect to science that will have "a clear and substantial impact on important public policies or private sector decisions," there will be circumstances where "access to data and methods cannot occur due to compelling interests such as privacy, trade secrets, intellectual property, and other confidentiality protections."[67] Significantly, the Guidelines do not instruct EPA to ignore such science. Rather, the Guidelines instruct that if underlying data or methods are unavailable, "EPA should, to the extent practicable, apply especially rigorous robustness checks to analytic results and carefully document all checks that were undertaken."[68] The Guidelines further explain: "Original and supporting data may not be subject to the high and specific degree of transparency provided for analytic results; however, EPA should apply, to the extent practicable, relevant Agency policies and procedures to achieve reproducibility, given ethical, feasibility, and confidentiality constraints."[69]

Far from instructing EPA not to consider scientific studies for which underlying data or models are unavailable, the EPA Information Quality Guidelines expressly acknowledge that EPA must balance a variety of important aims to fulfill its statutory obligations to protect public health and the environment. EPA explains in the guidelines that "most environmental statutes obligate EPA to act to prevent adverse environmental and human health impacts" and that "[f]or many of the risks that we must address, data are sparse and consensus about assumptions is rare."[70] Thus, rather than set rigid rules regarding what science and information EPA can rely upon in its rulemakings, EPA "seek[s] to strike a balance among fairness, accuracy, and efficient implementation."[71] EPA states: "Refusing to act until data quality improves can result in substantial harm to human health, safety, and the environment."[72]

As discussed *infra* at Section I.B.3.b)ii, even this Administration, in the context of promulgating regulations under TSCA, has adopted a regulatory definition of "best available

---

[65] *See* EPA Science Policy Council, *A Summary of General Assessment Factors for Evaluating the Quality of Scientific and Technical Information*.
[66] EPA, *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the EPA* (2002), https://www.epa.gov/quality/guidelines-ensuring-and-maximizing-quality-objectivity-utility-and-integrity-information.
[67] *Id.* at 21.
[68] *Id.*
[69] *Id.*
[70] *Id.* at 52.
[71] *Id.*
[72] *Id.*

science" expressly incorporating a multi-factor analysis, and that definition recognizes that public unavailability of data does not render a study incapable of being "best available science."

c) The courts

As EPA acknowledges in footnote 3 of the Proposal, in at least two instances the D.C. Circuit Court of Appeals has recognized that studies for which underlying data is not publicly available may constitute "best available science."[73] The D.C. Circuit's decisions in these cases further demonstrate that the public unavailability of a study's underlying data does not render a study incapable of constituting "best available science" otherwise unworthy of EPA's consideration.

In *American Trucking Associations* v. *EPA,* the petitioner challenged EPA's reliance on scientific studies for which underlying data was not publicly available in deciding to strengthen the national ambient air quality standards for particulate matter.[74] The Court held that the Clean Air Act did not require EPA to make public underlying data where EPA relied on the study itself and not the raw data underlying the study. The Court agreed with EPA's position that requiring agencies to obtain and publicize the data underlying all studies on which they rely "would be impractical and unnecessary."[75] Importantly, the Court concluded that:

> If EPA and other governmental agencies could not rely on published studies without conducting an independent analysis of the enormous volume of raw data underlying them, *then much plainly relevant scientific information would become unavailable to EPA for use in setting standards to protect public health and the environment*. . . . Such data are often the property of scientific investigators and are often not readily available because of. . .proprietary interests. . . or because of [confidentiality] arrangements [with study participants].[76]

The court accordingly recognized that ignoring relevant scientific information simply because the underlying data is not available would violate EPA's obligations to consider "best available science." *Coalition of Battery Recyclers Association v. EPA* involved another challenge to EPA's reliance on a scientific study for which the underlying data was not publicly available.[77] In that case, EPA had relied upon the study in question to determine the "concentration-response relationship between blood lead levels and IQ changes."[78] The D.C. Circuit again upheld EPA's reliance on studies without making the underlying data publicly available and explained, "raw data often is unavailable due to proprietary interests of a study's scientific investigators or confidentiality agreements with study participants."[79] Likewise, in *City of Waukesha v. EPA* the

---

[73] 83 Fed. Reg. at 18769.
[74] 283 F.3d 355, 372 (D.C. Cir. 2002).
[75] *Id.* at 372 (quoting National Ambient Air Quality Standards for Particulate Matter, 62 Fed. Reg. 38,652, 38,689 (July 18, 1997).
[76] *Id.* (emphasis added).
[77] 604 F.3d 613, 622-23 (D.C. Cir. 2010).
[78] *Id.* at 622.
[79] *Id.* at 623.

D.C. Circuit concluded that agency peer review satisfies the requirement to use best, peer-reviewed science and supporting studies.[80]

d) The Proposal

Finally, even the Proposal appears to concede that studies for which data is not publicly available could constitute the "best available science" that EPA is statutorily required to consider. The proposed exemption provision in section 30.9 makes it clear that EPA does not consider a study to be invalid or unsuitable for EPA's consideration based only on the public unavailability of underlying data or models. Specifically section 30.9 would give the Administrator discretion to authorize consideration of a scientific study where "[i]t is not feasible to ensure that all dose response data and models underlying pivotal regulatory science is publicly available." Of course, EPA could not have intended for proposed section 30.9 to provide the Administrator with discretion to take a study that is not "best available science" into consideration when promulgating a rulemaking. If the Administrator has discretion to allow consideration of a study for which it is infeasible to make the study's underlying data and models publicly available, then it obviously is not necessary for such underlying data and models to be publicly available for a scientific study to constitute "best available science." Yet, unless the Administrator elects to exercise his discretion under proposed section 30.9 and find that it is "infeasible" to make a study's underlying data and models publicly available, proposed section 30.5 broadly prohibits EPA from relying on the study in support of "significant regulatory actions."

Moreover, while proposed section 30.5's prohibition would apply to "pivotal regulatory science" used for "significant regulatory actions," the proposed rule says nothing to prohibit EPA's reliance on these studies for other agency purposes, such as in permitting, enforcement, or regulatory actions that do not qualify as "significant." Thus, EPA clearly does not believe that a study cannot be "best available science" based solely on the fact that underlying data and models are not publicly available.

In sum, if finalized, EPA's proposed rule would restrict EPA's ability to consider "best available science" when undertaking significant rulemakings, contrary to the numerous statutory directives discussed in detail below.

3. By prohibiting EPA from considering all valid and relevant studies when undertaking significant rulemakings, the proposed rule would prevent EPA from complying with an array of statutory provisions governing EPA's consideration of available science.

a) The Proposal Contravenes the Clean Air Act

[80] 320 F.3d 228, 247 (D.C. Cir. 2003).

Under Clean Air Act section 108(a),[81] EPA must establish air quality criteria for each air pollutant that serves as the basis for setting the national ambient air quality standards. Such criteria "shall accurately reflect the latest scientific knowledge useful in indicating the kind and extent of all identifiable effects on public health or welfare which may be expected from the presence of such pollutant in the ambient air, in varying quantities."[82] As explained above, the scientific community, EPA, and the courts have all concluded that lack of public availability of underlying data does not render the study invalid. And, consideration of such studies can be essential for EPA to fulfill Clean Air Act section 108(a)'s directive that it consider "the latest scientific knowledge" in establishing air quality criteria, that it consider studies "useful" in indicating effects of pollutants on ambient air, and in providing an adequate margin of safety in the standard itself. [83] Thus, EPA's proposal to bar EPA from considering such studies would prevent EPA from complying with its statutory obligation under Clean Air Act section 108(a).

Section 108(a)(2) says nothing about excluding information—its evident purpose is to be inclusive as to information to be considered. EPA's historic practice reflects this broad directive: each NAAQS review evaluates virtually all studies in the area, excluding none, but assigning appropriate weight based on study-by-study evaluation. Since the NAAQS provisions were enacted in 1970, EPA has conducted many NAAQS rulemakings. The agency does not establish *per se, a priori* rules regarding study inclusion or exclusion, but rather evaluates each of the individual studies—and there are thousands typically evaluated for each NAAQS review—on their merits based on reasoned criteria. While details of the development and review of the criteria and standards have evolved over time, in practice, EPA has endeavored to include all relevant scientific studies in the process, even providing provisional assessments of relevant literature that appears after the formal scientific review has been completed. Over the years, tens of thousands of peer-reviewed studies of health effects, exposure, and atmospheric interactions, and monitoring have been included in reviews of criteria and standards. A requirement that they must be excluded from consideration unless the raw data and full methodologies are made available for all of them is inconsistent with the legislative mandate and EPA's practice over the last 40 years.

Thus, a science regulation that applies to the NAAQS is unlawful unless EPA can show that the new standard can be established and implemented consistent with the applicable statutory requirements. To do so, EPA must prove that public unavailability of data means that a study does not constitute "latest scientific knowledge useful" in indicating effects on human health or welfare.[84] EPA's Proposal neither acknowledges this requirement nor explains how the Proposal would not violate this statutory command.

---

[81] 42 U.S.C. § 7408(a).
[82] 42 U.S.C. § 7408(a)(2).
[83] *Id.*
[84] 42 U.S.C. § 7408(a)(2).

For example, in past NAAQS reviews, EPA has considered the Harvard Six Cities study[85] and American Cancer Society studies[86], despite the fact that the data underlying these studies is not publicly available. These studies, however, are plainly "useful in indicating the kind and extent of all identifiable effects on public health or welfare."[87] These seminal studies have been part of the air quality criteria since the mid-1990s—they have thus been accepted as "useful" by separate panels of CASAC, and by EPA, in three separate NAAQS reviews. Their use has been upheld by the D.C. Circuit.[88] Both studies have been reanalyzed and validated by highly competent third-party reviewers (the Health Effects Institute) with access to the underlying data.[89] The study results have been reproduced many times over. [90] Extended follow-up analyses of the ACS and Harvard Six Cities studies provide consistent and stronger evidence of an association with PM 2.5 and mortality at even lower air quality distributions than had previously been observed.[91] This type of cumulative weight of evidence is highly probative in assessing both causality and in establishing the level of the NAAQS.[92] The proposal says almost nothing about any of these other attributes that not only make these studies "useful," but indeed make them particularly high quality and reliable.

The primary ozone NAAQS provides further examples of the pernicious effects the proposal would have. Among the key controlled human exposure studies demonstrating that exposure to ozone causes adverse health effects in even healthy subjects at levels below the level of the then-current NAAQS are Adams (2006) and Schelegle (2009).[93] These studies were sponsored by the American Petroleum Institute, which controls access to the underlying data. The American Petroleum Institute refused an EPA researcher access to the data of a related

---

[85] Dockery, D.W., Pope, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris Jr, B.G. and Speizer, F.E., *An association between air pollution and mortality in six US cities*, 329(24) New England Journal of Medicine 1753-1759 (1993).

[86] Pope, C.A., Thun, M.J., Namboodiri, M.M., Dockery, D.W., Evans, J.S., Speizer, F.E. and Heath, C.W., *Particulate air pollution as a predictor of mortality in a prospective study of US adults*, 151(3) American Journal of Respiratory and Critical Care Medicine 669-674 (1995); Krewski, D., Jerrett, M., Burnett, R.T., Ma, R., Hughes, E., Shi, Y., Turner, M.C., Pope, C.A. III, Thurston, G., Calle, E.E., Thun, M.J., *Extended Follow-up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality*, 140 Health Effects Institute, Boston, MA (2009).

[87] CAA section 108 (a)(2), 42 U.S.C. §7408(a)(2).

[88] *Coalition of Battery Recyclers Ass'n v. EPA*, 604 F.3d at 623.

[89] Krewski, Daniel, et al., *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality*, Health Effects Institute, Cambridge, MA (2000).

[90] *See* EPA, NCEA, *Integrated Science Assessment for Particulate Matter* (EPA/600/R-08/139F), 7-86 (2009).

[91] *See* EPA, *Policy Assessment for the Review of the Particulate Matter National Ambient Air Quality Standard* (EPA 452/R-11-003), 2-31 to 33 (Apr. 2011). *See also* Memorandum by Alison Cullen, Chair, SAB Work Group on EPA Planned Actions for SAB Consideration of the Underlying Science at 4 (May 12, 2018) (noting that "additional studies have confirmed the basic findings" of the Six Cities and American Cancer Society studies and that "the rigorous form of peer review and independent reanalysis" applied "has accomplished a measure of confidence in findings without public access to data and analytic methods.").

[92] *State of Mississippi v. EPA*, 744 F.3d 1334, 1344 (D.C. Cir. 2013) (endorsing EPA's weight of evidence approach, and stating that "incremental (and arguably duplicative) studies are valuable precisely because they confirm or quality previous findings or otherwise decrease uncertainty").

[93] *See* EPA, *Policy Assessment for the Review of the Ozone National Ambient Air Quality Standards* (EPA -452/R-14-006, 3-27, 4-10 (Aug. 2014).

Adams study it sponsored (Adams (1998)).[94] So not only would these evidently "useful" (under CAA section 108(a)(1)) studies be barred from consideration under the Proposal, but the Proposal creates a perverse incentive for industry to refuse access to study data. The published studies— peer reviewed—would obviously be providing information "useful" in indicating effects of air pollution, but the Proposal would not only bar their consideration but create an incentive for industry never to provide underlying data for any industry-sponsored study with a result not to industry's liking.

The most recent premiere long-term cohort study for PM is Domenici (2017) which found even greater effects of fine particles at levels below EPA's current standards.[95] This study used a Medicare database available to any research group that can guarantee confidentiality of personal data.[96] Yet the proposal could evidently bar consideration of this powerful study.[97]

NAAQS must be requisite to protect the public health, and to provide an "adequate margin of safety" in doing so.[98] The proposal violates this central statutory requirement. NAAQS are required to provide this margin of safety "to build a buffer to protect against uncertain and unknown dangers to human health."[99] EPA's Proposal would build a buffer against using the very studies necessary to guard against these dangers.[100]

b) EPA's Proposal contravenes the Toxic Substances Control Act (TSCA).

i. TSCA expressly requires that EPA consider reasonably available information and EPA's proposal would preclude EPA from considering some reasonably available information.

When Congress amended TSCA through passage of the Frank R. Lautenberg Chemical Safety for the 21st Century Act (Lautenberg Act), Congress provided a number of detailed instructions on how EPA should consider scientific information with respect to chemical substances; EPA's proposal contradicts Congress's carefully crafted scheme. In particular, Congress included a provision specifically requiring that EPA consider all "reasonably available

---

[94] *See* EPA, *First External Review Draft Integrated Science Assessment for Ozone and Related Photochemical Oxidants* (EPA/600/R-10/076A), 6-7 n. 1 (Feb. 2011).

[95] Qian Di et. al., *Air Pollution and Mortality in the Medicare Population*, 376 New England Journal of Medicine 2513 (2017), https://www.nejm.org/doi/pdf/10.1056/NEJMoa1702747.

[96] *See* CMS, *Limited Data Set (LDS) Files*, https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/DUA_-_NewLDS.html (last accessed Aug. 9, 2018) (noting data requires a signed data use agreement and data cannot be disclosed).

[97] *See* 83 Fed. Reg. 18768, 18773, Proposed section 30.5 final sentence ("where data is controlled by third parties, EPA shall work with those parties to endeavor to make the data available in a manner that complies with this section"). There appears to be some interaction required before third party studies are considered to be publicly available.

[98] CAA section 109(b); 42 U.S.C. § 7409(b).

[99] *State of Mississippi,* 744 F.3d at 1353.

[100] *See American Farm Bureau v. EPA*, 559 F.3d 512, 525-26 (D.C. Cir. 2009) (remanding primary Particulate Matter NAAQS because inadequate consideration of certain epidemiologic studies resulted in a standard lacking an adequate margin of safety).

information."[101] When making decisions about testing or the risk evaluation or regulation of new or existing chemicals, "the Administrator shall take into consideration *information* relating to a chemical substance or mixture, including hazard and exposure information, under the conditions of use, *that is reasonably available to the Administrator*." 15 U.S.C. § 2625(k) (emphases added). But under EPA's proposed rule, EPA would often be precluded from considering such reasonably available information if all the underlying data and models were not publicly available. *See* 83 Fed. Reg. at 18,769 n.3 (stating that proposal "would preclude [EPA] from using [non-public] data in future regulatory actions"). EPA's proposal violates the plain language of TSCA § 26(k), as well as Congress's clear purpose of ensuring that EPA consider all reasonably available information relating to a chemical when making a decision about the chemical.

Under its plain language, "available" means "able to be used or obtained; at someone's disposal."[102] Congress chose this standard to ensure that EPA would make decisions based on all reasonably available information. S. Rep. No. 114-67 at 9 (June 18, 2015) ("The section … requires EPA to consider reasonably available information about potential hazards and exposures of a chemical substance under the conditions of use when making decisions under TSCA.… The Committee intends that EPA systematically search for and identify relevant information that is available to inform safety assessments and determinations."); Oversight of the Environmental Protection Agency's Progress in Implementing Inspector General and Government Accountability Office Recommendations: Hearing before the Subcomm. on Superfund, Waste Management, and Regulatory Oversight of the S. Comm. on Environment and Public Works, 114th Cong. at 63 (June 14, 2016) ("[F]or the EPA to properly evaluate and regulate toxic substances, it is essential that they have the most up-to-date chemical and toxicity data available."). Congress also selected this standard to avoid paralysis by analysis—Congress wanted EPA to act on available information and not to postpone action waiting for new or perfect information to become available. *See, e.g.*, 162 Cong. Rec. S3511, S3517 (daily ed. June 7, 2016) (referring to "information reasonably available to EPA" as "ensur[ing] that such considerations do not require additional information to be collected or developed"). "Congress recognized the need to use available studies, reports and recommendations for purposes of chemical assessments rather than creating them from whole cloth." *Id.* at S3522. And Congress intended for EPA to consider studies even when they had not undergone all possible forms of vetting. "[I]n instances where there were other studies and reports unavailable at the time of the [National Academy of Sciences] recommendations, EPA should take advantage of those studies and reports in order to ensure that the science used for chemical assessments is the best available and most current science." *Id.* at S3522. Congress intended for EPA to consider all reasonably available information, and EPA's proposal would thwart that clear purpose.

Notably, this Administration has adopted two regulations under the amended TSCA defining reasonably available information. These regulations generally provide that:

> Reasonably available information means information that EPA possesses or can reasonably generate, obtain, and synthesize for use in risk evaluations, considering the deadlines specified in TSCA [for action]. Information that meets the terms of the

[101] Pub. L. No. 114-182, § 17(k), 130 Stat. 448, 502 (June 22, 2016) (codified at 15 U.S.C. § 2625(k)).
[102] *Oxford American Dictionary* 111 (3d ed. 2010).

> preceding sentence is reasonably available information whether or not the information is confidential business information, that is protected from public disclosure under TSCA section 14.

40 C.F.R. § 702.33; *see also* 40 C.F.R. § 702.3 (similar definition for prioritization decisions). This bears no resemblance to the limitations put forward in the Proposal. Indeed, EPA has defined "reasonably available information" to include information EPA withholds as Confidential Business Information (CBI) under TSCA § 14. 15 U.S.C. § 2613. If the proposed rule forecloses EPA from considering information that cannot be fully disclosed, as it appears to do, then EPA cannot comply with both these regulations and the proposed rule.

EPA's proposal also violates other provisions of TSCA that expressly require EPA to act on "available information." For example, in preparing risk evaluations for existing chemicals, EPA "shall integrate and assess *available information* on hazards and exposures for the conditions of use of the chemical substance, including information that is relevant to specific risks of injury to health or the environment and information on potentially exposed or susceptible subpopulations identified as relevant by the Administrator."[103] Under the proposed rule, EPA would not be able to integrate and assess available information where all underlying data has not been disclosed. Similarly, when developing regulations for existing chemicals, EPA "shall consider and publish a statement based on *reasonably available information* with respect to" a number of factors, including the effects of the chemical on health and the environment.[104] But under the proposed rule, EPA cannot consider all reasonably available information when assessing those health and environmental effects.

Indeed, TSCA § 4(f) imposes a duty upon EPA to initiate regulation in response to any available information that meets certain substantive standards. However, if all the underlying information were not available, EPA's proposed rule would then foreclose EPA from considering that information during the resulting rulemaking. Congress would not have created a scheme where EPA *must* act in response to certain information but then cannot consider that information in taking action. Specifically, under TSCA § 4(f):

> Upon the receipt of—(1) *any information* required to be submitted under this Act, or (2) *any other information available* to the Administrator—which indicates to the Administrator that there may be a reasonable basis to conclude that a chemical substance or mixture presents a significant risk of serious or widespread harm to human beings, the Administrator shall, … initiate applicable action under section 5, 6, or 7 to prevent or reduce to a sufficient extent such risk or publish in the Federal Register a finding, made without consideration of costs or other nonrisk factors, that such risk is not unreasonable.[105]

Thus if "any … information available" to EPA provides a reasonable basis to conclude that a chemical "presents a significant risk of serious or widespread harm to human beings," then EPA must initiate action to regulate the chemical. But under EPA's proposed rule, EPA would then be

---

[103] 15 U.S.C. § 2605(b)(4)(F)(i) (emphasis added).
[104] *Id.* § 2605(c)(2)(A) (emphasis added).
[105] 15 U.S.C. § 2603(f) (emphases added).

required to ignore the information triggering this duty when crafting the final regulation unless the source of the information fully disclosed all underlying data. That result clearly contradicts Congress's intent, which was to create a duty for EPA to react to any available information meeting the substantive standard of TSCA § 4(f).

In sum, Congress repeatedly directed EPA to consider all reasonably available information when making decisions under TSCA. The proposed rule would illegally preclude EPA from considering available information. The two cannot be reconciled, and the rule is unlawful.

ii. TSCA requires an agency to act on the "best available science," meaning that EPA must consider all available science and assess the quality of the science based on a variety of factors.

EPA's proposed blanket prohibition against basing a rulemaking on science for which underlying data or models are not publicly available would be particularly hard to reconcile with the "best available science" standard as articulated in TSCA, which clearly contemplates a case-by-case analysis in which EPA weighs a variety of factors when identifying the best available science. The relevant provision of TSCA requires that:

> (h) Scientific standards. In carrying out sections 4, 5, and 6, to the extent that the Administrator makes a decision based on science, the Administrator shall use scientific information, technical procedures, measures, methods, protocols, methodologies, or models, employed in a manner consistent with the *best available science*, and shall consider *as applicable*—
> (1) the extent to which the scientific information, technical procedures, measures, methods, protocols, methodologies, or models employed to generate the information are reasonable for and consistent with the intended use of the information;
> (2) the extent to which the information is relevant for the Administrator's use in making a decision about a chemical substance or mixture;
> (3) *the degree of clarity and completeness* with which the data, assumptions, methods, quality assurance, and analyses employed to generate the information *are documented*;
> (4) the extent to which the variability and uncertainty in the information, or in the procedures, measures, methods, protocols, methodologies, or models, are evaluated and characterized; and
> (5) *the extent of independent verification or peer review* of the information or of the procedures, measures, methods, protocols, methodologies, or models.[106]

Thus, Congress provided EPA with factors to guide its consideration of the "best available science," and Congress did not make the public disclosure of all underlying data a requirement for material to be the "best available science." Quite the opposite; Congress included aspects of disclosure and independent review as parts of factors to be considered when weighing scientific information. But these are just aspects of five different factors to be weighed "as applicable," and

[106] 15 U.S.C. § 2625(h) (emphases added).

Congress clearly contemplated that EPA would sometimes rely on science that does not meet the proposed rule's requirement of full disclosure of all underlying data.

First, Congress directed EPA to consider these factors when weighing particular information; Congress specifically did not develop (or direct EPA to develop) bright-line criteria for eliminating information from consideration entirely. Thus, each factor includes the phrase "degree of" or "extent to which," without identifying any threshold that would be disqualifying.[107] This shows that Congress intended these factors to help EPA assess the weight information should be given based on its relative scientific reliability, not to create minimum thresholds of reliability below which information must be ignored by EPA altogether. For EPA to insert a screen on top of these factors—excluding information where the underlying data and models are not publicly available as required by the proposed rule—contradicts Congress's unambiguous intent about how EPA should approach its assessment of the best available science.

Second, Congress made the "degree of clarity and completeness" with which the underlying data is documented to be part of one factor for EPA to consider in evaluating whether a particular study is the "best available science."[108] But EPA must also consider "the degree of clarity and completeness" with which "assumptions, methods, quality assurance, and analyses" are documented as well.[109] Thus, Congress contemplated that EPA would still rely on some studies that did *not* document completely all the underlying data, much less disclose all of that information.

Third, Congress made "the extent of independent verification *or* peer review of the information *or* of the procedures, measures, methods, protocols, methodologies, or models" another factor to be weighed when considering whether information is the "best available."[110] Notably, Congress's choice of the disjunctive "or" reflects that "peer review" can be an adequate alternative to "independent verification," and Congress did not require that either "independent verification *or* peer review" be accomplished through public availability of data as required in the proposed rule. Moreover, Congress contemplated scenarios where EPA would give more weight to evidence even if the "information" had not undergone "independent verification or peer review" based on the extent to which the "procedures, measures, methods, protocols, methodologies, or models" had done so.

Fourth and most importantly, EPA cannot rationally elevate the interest in public disclosure of all underlying data above all the other factors that Congress expressly required EPA to consider in evaluating science. Congress required EPA to consider these five factors "as applicable" when weighing information, and Congress did not make full public availability of underlying data one of the factors, much less a decisive or absolute one.

---

[107] *See, e.g.* 15 U.S.C. § 2625(h)(1) ("*the extent to which* the scientific information…[are] consistent with the intended use of the information") (emphasis added).
[108] 15 U.S.C. § 2625(h)(3).
[109] *Id.*
[110] 15 U.S.C. § 2625(h)(5).

This administration recently adopted a regulatory definition of "best available science" for purposes of TSCA which expressly incorporated consideration of these five factors and was otherwise inspired by use of the term in the Safe Drinking Water Act (SDWA).[111] EPA defined the phrase:

> Best available science means science that is reliable and unbiased. Use of best available science involves the use of supporting studies conducted in accordance with sound and objective science practices, including, when available, peer reviewed science and supporting studies and data collected by accepted methods or best available methods (if the reliability of the method and the nature of the decision justifies use of the data). Additionally, EPA will consider as applicable:
>
> [TSCA § 26(h)(1)(5) factors][112]

According to EPA in selecting this definition, "the Agency is remaining consistent with the current approach already used Agency-wide, while also acknowledging the specific standards under TSCA."[113] Notably, this definition does not require public disclosure of all underlying data for science to be the "best available science," yet many studies that meet this definition of "best available science" would be excluded under EPA's proposed rule.

EPA's Proposal cannot be reconciled with EPA's existing definition of best available science, with decades of court and agency precedent, or with text of the statute. When a statute requires the agency to make a decision based on the "best available science," it would be unlawful to follow EPA's proposed rule.

iii. EPA's proposed rule also contradicts TSCA's requirement that decisions be made based on the weight of the scientific evidence.

TSCA § 26(*i*) requires EPA to make decisions regarding testing and regulating new and existing chemicals "based on the weight of the scientific evidence."[114] If EPA excludes certain information, as proposed, then EPA will not be able to weigh the evidence as a whole.

Indeed, this administration recently adopted a regulation defining "weight of scientific evidence" to mean "a systematic review method … that uses a pre-established protocol to *comprehensively*, objectively, transparently, and consistently, identify and evaluate *each stream of evidence*, including strengths, limitations, and relevance of *each* study and to integrate evidence as necessary and appropriate based upon strengths, limitations, and relevance."[115] Systematic reviews consider the entire body of scientific evidence, but EPA's proposed rule would prevent EPA from conducting true systematic review because it would prohibit the Agency from considering studies where the data were not publicly available and it would

---

[111] *See* 82 Fed. Reg. 33,726, 33,731 (July 20, 2017), 42 U.S.C. § 300g-1(b)(3)(A).
[112] 40 C.F.R. § 702.33.
[113] 82 Fed. Reg. at 33,731.
[114] 15 U.S.C. § 2625(*i*).
[115] 40 C.F.R. § 702.33 (emphases added).

eliminate studies based on criteria other than their "strengths, limitations, and relevance."[116] If the proposed rule forecloses EPA from considering information that cannot be fully disclosed, as it appears to do, then EPA cannot comply with this regulation and the proposed rule.

In sum, EPA's proposed rule is inconsistent with TSCA's plain text. EPA should not adopt the proposed rule because it cannot be reconciled with the agency's duties under TSCA.

iv. Section 10 of TSCA does not authorize this proposal.

Nothing in Toxic Substances Control Act (TSCA) § 10 authorizes EPA to exclude scientific information during rulemakings on any basis. Section 10 authorizes EPA to research and develop information for purposes of carrying out TSCA.[117] Section 10 also authorizes EPA to develop systems to collect and disseminate information about chemical substances.[118] But TSCA § 10 is silent regarding rulemaking or EPA's use of scientific information in rulemaking. It does not authorize EPA to exclude scientific information on *any* basis; if anything, TSCA § 10 reflects a congressional judgment that EPA should be prepared to use any and all "toxicological and other scientific information which could be useful to the Administrator in carrying out the purposes of this [Act]."[119]

c) EPA's Proposal contravenes the Safe Drinking Water Act.

The Safe Drinking Water Act requires EPA to issue national drinking water regulations setting required purity levels for water from public water supply systems.[120] Before regulating, the Administrator must conclude that the contaminant at issue "may have" an adverse effect on the health of persons.[121] In regulating, the Administrator must consider "the best available public health information"[122] The section adds that in setting regulations, the Administrator "shall use ...the best available, peer-reviewed science and supporting studies conducted in accordance with sound and objective scientific practices" and in addition "data collected by accepted methods or best available methods."[123] When Congress promulgated these statutory requirements in 1996, the Senate Committee on Environment and Public Works[124] explained that the "Administrator has a *duty* to seek and rely upon the best available science and information to support.... [m]any

---

[116] *Id.*
[117] *See* 15 U.S.C. § 2609(a) ("The Administrator shall ... conduct such research, development, and monitoring as is necessary to carry out the purposes of this [Act]."); *see also* 15 U.S.C. § 2609(c), (d), (e).
[118] *See* 15 U.S.C. § 2609(b), (c), (g).
[119] 15 U.S.C. § 2609(b)(2)(A).
[120] 42 U.S.C. § 300g-1.
[121] *Id.* at (b)(1)(A)(i).
[122] *Id.* at (b)(1)(B)(ii)(II).
[123] 42 U.S.C. § 300g-1(b)(3)(A). *See City of Waukesha v. EPA*, 320 F.3d at 247-48 (D.C. Cir. 2003) (holding that agency peer review satisfies requirement to use best, peer-reviewed science and supporting studies); *City of Portland v. EPA*, 507 F 3d 706, 716 (D.C. Cir. 2002) (same).
[124] The Report of the Senate Committee on Environment and Public Works is authoritative on these provisions, as the language adopted in the Committee bill (S.1316) on the use of science was adopted verbatim in Pub. L. 104-182. *See* S. Rep. 104-169 at p. 121 and Pub. L. 104-182 at §103.

of the most important activities including selecting contaminants for regulation, setting standards, designing analytical methods and structuring waivers, variances and exemptions."[125]

By restricting EPA to considering only those scientific studies for which underlying data, models, and other information is publicly available, EPA's proposal prevents EPA from complying with the SDWA directive that it consider the "best available" public health information and science when setting SDWA standards. Specifically, as explained above, the public will not necessarily have access to the underlying information used to produce the "best available, peer-reviewed science and supporting studies."[126] Nowhere does the SDWA authorize EPA to ignore such studies based on the public unavailability of underlying information. Thus, regardless of the merits of the core objective of EPA's proposal—"to ensure that the regulatory science underlying its actions is publicly available in a manner sufficient for independent validation" (proposed § 30.1 "What is the purpose of this subpart?"), EPA's attempt to elevate this objective above the agency's statutory obligation to consider the "best available" science when promulgating SDWA standards is unlawful.[127]

4. EPA's proposed exemption provision does not remedy the unlawfulness of prohibiting EPA from considering valid and relevant studies due to the public unavailability of underlying data and methods.

Though the proposed exemption provision in section 30.9 would grant the EPA Administrator discretion to authorize the agency to consider studies for which underlying data or models are not publicly available, this provision is insufficient to remedy the proposed rule's unlawfulness and detrimental impacts. It is well established that existence of a waiver or exemption mechanism cannot be used to justify a provision otherwise beyond an agency's legal authority. *Dimension Financial Corp. v. Board of Governors of Federal Reserve System*, 744 F.2d 1402, 1410 (10th Cir. 1984) ("The possible exception to the initial impact of Regulation Y (Part 225.21(B)(4)) contains requirements with no objective standard and thus unbounded agency discretion. This as a device to meet objections to the new regulation cannot cure the exercise of powers denied by Congress or not provided for by Congress. *Public Utilities Comm. of Calif. v. United States,* 355 U.S. 534 (1958); *In re Surface Mining Regulation Litigation,* 627 F.2d 1346 (D.C. Cir. 1980); *ALLTEL Corp. v. FCC*, 838 F.2d 551, 561 (D.C. Cir. 1988) ("The FCC cannot save an irrational rule by tacking on a waiver procedure. 'The very essence of waiver is the assumed validity of the general rule . . . .')(citing *WAIT Radio v. FCC*, 418 F.2d 1153, 1158 (D.C. Cir. 1969)); *United States Telecom Ass'n v. FCC*, 359 F.3d 554, 571 (D.C. Cir. 2004) ("Moreover, even if the FCC had adopted some lawful mechanism for making exemptions from its general national rule, it could not necessarily rely on the existence of that mechanism as the sole justification for not adopting a more narrowly tailored rule. . . . [T]he mere existence of a safety valve does not cure an irrational rule.")

[125] S. Rep. 104-169 at 28 (emphasis added).
[126] 42 U.S.C. § 300g-1(b)(3)(A).
[127] 83 Fed. Reg. at 18773.

First, while the statutory provisions described above *require* EPA to consider best available science and other relevant information when making regulatory decisions, *see, e.g.*, Safe Drinking Water Act, 42 U.S.C Section 300g-1(b)(3)(A)(i) ("The Administrator *shall* use the best available, peer reviewed science."), the Administrator has discretion over whether to grant an exception. *See* Proposed § 30.9 ("The Administrator *may* grant an exemption to this subpart on a case-by-case basis…")(emphasis added).[128] Where a statute requires that the agency consider certain information in reaching a decision, EPA cannot promulgate a rule that gives the Administrator discretion over whether to allow such consideration.

Second, the only basis on which the Administrator may grant an exemption under Proposed § 30.9 is that it "is not feasible" to "ensure that all dose response data and models underlying pivotal regulatory science is publicly available" as the rule requires.[129] However, the Proposal does not explain how "feasibility" is to be determined in this context—or even whether the term encompasses practical feasibility, cost-effectiveness, or other considerations. Moreover, there can easily be situations where it is theoretically "feasible" to make underlying data publicly available, but this information is nonetheless not publicly available. For example, a scientist who intends to rely on the same data to publish multiple papers may be disinclined to make that data available to competitors.[130] Yet, because it is technically "feasible" to make the underlying data publicly available, the proposed rule would not even provide the Administrator with authority to grant an exemption authorizing such consideration, thus forcing the Administrator to violate the law.

Third, even if it were lawful for EPA to ignore relevant science, the exemption provision is arbitrary, as it does not define sufficient criteria or process steps by which the Administrator may decide to exempt a study. The provision instructs the Administrator to rely on a handful of broad (and highly manipulable) policy considerations in determining whether it would be infeasible to make data and methods publicly available.[131] These factors could be applied broadly to give the Administrator nearly absolute discretion. From the face of the Proposal, it is not even clear that the Administrator would be required to provide a public, written explanation of his decision to grant (or deny) a waiver. This lack of accountability could lead to the arbitrary exclusion of studies the Administrator unilaterally chooses to not exempt.

---

[128] 83 Fed. Reg. at 18774.

[129] 83 Fed. Reg. at 18774.

[130] Or in cases where companies jointly funded research it may be unclear who owns the data and has the right to share it, and companies may be reluctant to share it with competitors. *See, e.g.,* National Academies of Sciences, Engineering, and Medicine, *Principles and obstacles for sharing data from environmental health research: Workshop summary*, 45 The National Academies Press (2016), https://www.nap.edu/catalog/21703/principles-and-obstacles-for-sharing-data-from-environmental-health-research. ("As you can imagine. . . not all competitors play nicely together. Some even resort to gamesmanship to try to exclude competitors from the market. Things can get nasty and messy in a hurry in these discussions.").

[131] *See* 83 Fed. Reg. at 18774. Under §30.9(a), the Administrator should consider whether it is infeasible "in a fashion that is consistent with law, protects privacy, confidentiality, confidential business information, and is sensitive to national and homeland security." §30.9(b) references 70 Fed. Reg. 2664, which exempts peer review in situations of "disseminations of sensitive information related to certain national security, foreign affairs, or negotiations involving international treaties and trade where compliance with this Bulletin would interfere with the need for secrecy or promptness."

Finally, the exemption provision is impractical and likely could not be implemented effectively. According to the Congressional Budget Office, EPA "relies on about 50,000 scientific studies annually to perform its mission," and at times, relies on thousands of studies for one action.[132] Many of the studies that would be affected by this rule are complex and include large datasets that would lead to an extensive decision-making process under the exemption provision. EPA does not include any rationale in the proposal justifying how the Administrator could reasonably decide to exempt studies on a case-by-case basis given the tens of thousands of studies EPA considers each year. This provision could create a large backlog, which would result in important studies being effectively removed from EPA consideration because of the need to finalize a regulation before an exemption for every relevant study is granted. Accordingly, the exemption provision fails to safeguard against the unlawful exclusion of valid science from EPA's regulatory process.

C. EPA's Proposed Rule Would Violate the Information Quality Act.

EPA's proposed rule is also unlawful because it exceeds EPA's authority under Section 515(a) of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 16-554; H.R. 5658), commonly referred to as the Information Quality Act.[133] Specifically, the Information Quality Act requires EPA promulgate data quality guidelines that are consistent with those promulgated by the Office of Management and Budget. Contrary to EPA's assertion in the preamble to the proposal, the Proposed Rule is not consistent with OMB's data quality regulations.

The OMB Guidelines recognize that data availability is not necessary to high quality science, but is one among many factors. While imposing high standards of quality, objectivity, utility, and integrity of information disseminated by Federal Agencies, the Guidelines recognize the need to implement controls "flexibly, and in a manner appropriate to the nature . . . of the information to be disseminated."[134] As part of ensuring "objectivity" these guidelines encourage agencies that disseminate influential scientific, financial, or statistical information, "to include a high degree of transparency about data and methods to facilitate the reproducibility of such information by qualified third parties."[135] However, they emphasize the need to treat certain data differently, due to privacy and confidentiality concerns.[136] In fact, the OMB Regulations specifically declare that "[w]ith regard to original and supporting data related thereto, *agency guidelines shall not require that all disseminated data be subjected to a reproducibility requirement.*"[137] Rather, the OMB Guidelines instruct that agencies "identify, in consultation with the relevant scientific and technical communities, those particular types of data that can

---

[132] Congressional Budget Office, *Cost Estimate: H.R. 1430* 2-3 (March 29, 2017), https://www.cbo.gov/system/files/115th-congress-2017-2018/costestimate/hr1430.pdf.
[133] Codified at 44 U.S.C. 3504(d)(1) and 3516.
[134] OMB's *Guidelines Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information*, 67 Fed. Reg. 8,452, 8,453 (Feb. 22, 2002).
[135] 67 Fed. Reg. at 8460.
[136] *OMB's Guidelines Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information*, 67 Fed. Reg. 8, 452, 8,460 (Feb. 22, 2002) (interest in making data publicly available "does not override other compelling interests such as privacy, trade secrets, intellectual property, and other confidentiality protections").
[137] 67 Fed. Reg. at 8460 (emphasis added).

practicable [sic] be subjected to a reproducibility requirement, given ethical, feasibility, or confidentiality constraints."[138] The OMB Regulations further explain that while "[m]aking the data and methods publicly available will assist in determining whether analytic results are reproducible...*the objectivity standard does not override other compelling interests such as privacy, trade secrets, intellectual property, and other confidentiality protections.*[139] OMB explains that "where public access to data and methods will not occur due to other compelling interests, agencies shall apply especially rigorous robustness checks to analytic results and document what checks were undertaken."[140]

By outright prohibiting EPA from relying on a study to support a significant rulemaking if that study's underlying data and models are not publicly available, EPA's proposed rule departs from OMB's unambiguous language instructing agencies that they "shall not" require that all data and models be subject to the reproducibility requirement, and that "the objectivity standard does not override other compelling interests."[141] The fact that EPA's proposed rule includes a discretionary "exemption" provision does not correct this problem, as that provision would not require the Administrator even to consider whether an exemption is warranted, let alone grant such an exemption under appropriate circumstances.

Because Congress expressly granted OMB the authority to set guidelines for data quality and instructed agencies like EPA to follow OMB's lead, EPA lacks statutory authority to adopt a regulation that is contrary to OMB's guidelines. Accordingly, EPA's proposed regulation violates the Information Quality Act and must be withdrawn.[142]

II. EPA's Proposed Rule is Unreasonable and Arbitrary and Capricious.

In addition to violating the requirements of the various statutes that EPA administers or is subject to, the Proposal suffers from a total failure to consider important dimensions of the profound shift in policy that it implements. In the Proposal, EPA neglects to consider the many legitimate reasons why a study's underlying data may not be publicly available—reasons that have nothing to do with the quality of the study—and fails to offer solutions consistent with these legitimate limitations. EPA makes vague gestures to various guidelines and practices issued by other agencies and scientific organizations, none of which actually support the Proposal's radical position that EPA should exclude consideration of studies that rely upon confidential data. EPA does not even establish that there is a real problem that the Proposal would actually address: nowhere in the Proposal does EPA identify any prior agency action that has been called into serious question due to a failure to release study data. EPA's utter failure "to consider an important aspect of the problem" and to provide an explanation for the Proposal

---

[138] 67 Fed. Reg. at 8460. There is no indication that EPA consulted with the scientific and technical community—or even its own Science Advisory Board—before proposing to require that the underlying data and models be made publicly available for all pivotal regulatory science regardless of ethical, feasibility, or confidentiality constraints.
[139] 67 Fed. Reg. at 8460 (emphasis added).
[140] 67 Fed. Reg. at 8460.
[141] *See* 67 Fed. Reg. at 8460.
[142] *Prime Time Int'l Co. v. Vilsack*, 599 F.3d 678, 685 (D.C. Cir. 2010) ("[B]ecause Congress delegated to OMB authority to develop binding guidelines implementing the IQA, we defer to OMB's reasonable construction of the statute.")

that is consistent with the evidence before the agency renders the Proposal wholly arbitrary and capricious. *See Motor Vehicle Mfrs. Ass'n. v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29, 43 (1983). Likewise, EPA's failure to explain its 180-degree change in position from its former belief that the lack of publicly-available data does not render a study inappropriate for consideration in regulating is a hallmark of arbitrary and capricious decision-making. *FCC v. Fox Telev. Stations, Inc.*, 556 U.S. 502, 515-16 (2009).

A. EPA Failed to Consider the Legitimate Reasons That Underlying Data May Not be Made Publicly Available, or to Propose Solutions to Remedy These Actual Limitations.

1. There are multiple reasons why underlying data are not publicly available for all studies.

There are legal and ethical requirements that restrict making public the data underlying studies, including rules to shield private personal information, requirements to maintain confidential business information, situations where obtaining the necessary permissions to release data are logistically difficult or impossible, and situations in which researchers have made significant investments in developing datasets that they intend to continue to work with for future studies. Not all of these barriers can be overcome, nor can they be overcome in every case. While there are ways potentially to address some of them, they can be extremely costly and burdensome, and/or may harm the prospects for further research. Accordingly, while the scientific community has made efforts to make more data publicly available, to the best of our knowledge all of the policies adopted by government and academic journals recognize that data is not, and need not be, publicly available to evaluate their quality.

a) Strong legal and ethical requirements limit the release of data in human subjects studies.

Particularly with respect to human subjects, there are strong legal and ethical privacy and confidentiality protections, which researchers are bound to respect.[143] In some cases, researchers would be subject to civil or criminal penalties for violations.[144]

The environmental health dose response studies targeted by EPA's proposal are likely to include human population studies (or epidemiological studies). Often the best available epidemiological studies contain extensive and sensitive data on individuals, such as environmental exposures, medical history (such as infant reproductive developmental abnormalities, children's behavioral and development problems, heart attacks or dementia among the elderly), dates of birth, residential address, drug use, race, socio-economic status (income,

---

[143] *See, e.g.*, The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, *The Belmont Report* (Apr. 18, 1979), https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf; *Federal Policy for the Protection of Human Subjects; Final Rule*, 82 Fed. Reg. 7,149 (Jan. 19, 2017); HIPAA Privacy Rule, 45 C.F.R. §§ 160, 164.102-06, 164.500-534.

[144] *See,* The Health Insurance Portability and Accountability Act of 1996 (HIPAA), Public Law 104-191 (enacted Aug. 21, 1996) (providing for criminal and civil penalties for violations).

education), status of subjects' marriages, employment history, etc. For example, air pollution studies commonly use residential address information to assign air pollution exposures and link them to health effects.[145] Other studies focused on genetically susceptible populations may also be linked to genetic databases or contain information on key genetic mutations that are strongly predictive of serious health risks, such as risk of Alzheimer's disease, and are thus very sensitive.[146]

To conduct these studies, investigators must obtain informed consent from the study participants to collect protected health information, and investigators must sign documents promising to protect the privacy of this individually identifiable health information. Absent complex, difficult and costly de-identification and redaction techniques, these data simply cannot be released publicly. As discussed below in section II.A.2.b), in some cases such techniques are simply not applicable or still leave significant risk of breach of privacy.

Additional protections apply to specific types of human subject information. For example, medical records are subject to strict requirements governing the use and disclosure of such information under the Health Insurance Portability and Accountability Act of 1996 (HIPAA).[147] HIPAA requires researchers to protect identifiable information, and it provides that such information may only be disclosed for research purposes with the written consent of the person providing the information.[148]

Another limitation on public availability of data is the requirement under the Federal Policy for the Protection of Human Subjects (also known as the Common Rule) that for all federally funded studies involving human research subjects, researchers must first obtain Institutional Review Board (IRB) approval and informed consent from study participants.[149]

An IRB reviews each human subjects research project to ensure that the specific research protocol protects individual rights. Participants must be notified about the degree to which the confidentiality of their records will be maintained, and must receive appropriate notification and

---

[145] *See, e.g.*, Kaufman, Joel D., et al., *Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study*, 388.10045 The Lancet 696-704 (2016).

[146] *See, e.g.*, Richardson JR, Roy A, Shalat SL, von Stein RT, Hossain MM, Buckley B, Gearing M, Levey AI, German DC, *Elevated serum pesticide levels and risk for Alzheimer disease*, 71(3) JAMA Neurology 284-90 (Mar. 1, 2014).

[147] Public Law 104 – 191.

[148] National Research Council, *Expanding Access to Research Data: Reconciling Risks and Opportunities*, The National Academies Press (2005).

[149] 45 C.F.R. §§ 46.101-124 is the U.S. Department of Health and Human Services ("HHS") citation for the Common Rule. A total of 18 federal agencies have adopted it; each agency has its own separate entry in the Code of Federal Regulations. This federal rule governs ethical constraints that federally funded studies must follow, including academic research, responding to earlier concerns of ethical lapses in medical research. *See, e.g.*, Jerry Menikoff, *Could Tuskegee happen Today?,* 1 St. Louis U. J. Health L. & Pol'y 311, 312-16 (2008) (describing the Congressional response to public outcry when the details of the Tuskegee experiment were brought to light). The thrust of the Common Rule is to address such matters of research ethics as informed consent, informational risk, and institutional oversight when research involves human subjects.

give consent if study data is to be shared outside the research team.[150] The IRB also considers risks to the participants and how use of the information obtained may adversely impact the rights and welfare of the subjects.[151] Most institutions have committed to comply with the Common Rule for all of their research, even when it is not federally-funded.[152]

For studies that had received IRB approval prior to finalization of this proposed rule, there may be no practical opportunity to make the data publicly available. Even for new studies going forward, it may be extremely difficult, require additional (often unavailable) funding for elaborate protective measures, or simply impossible to obtain IRB approval for protocols that would allow the data to be made publicly available.

EPA's own Science Advisory Board voiced these concerns that EPA was discounting the challenges to making even limited releases of data, saying:

> The proposed rule oversimplifies the argument that "concerns about access to confidential or private information can, in many case, be addressed through the application of solutions commonly in use across some parts of the Federal government." For studies already completed or underway, the participation of human subjects is undertaken according to terms approved by the cognizant IRB. These terms can vary from study to study. In some cases, the data cannot be released simply by redacting portions of it. For example, data may have been collected with an assurance to the participating individuals that their data would be kept confidential.[153]

Some researchers might respond by choosing to work only on public administrative datasets, but this would harm rather than strengthen science quality by curtailing scientific inquiry. Thus, the effects of EPA's proposed approach would cause some researchers to choose not to pursue research with human subjects, stifling scientific discovery, while others would forgo compliance with EPA's regulatory requirements and have their research ignored by EPA. As a result, EPA's proposal would both discourage the development of best available science as well as EPA's use of it.

> b) There are especially significant barriers to public release of underlying data and models from studies that have already been completed.

With respect to studies that have already been completed, there are additional formidable barriers to public release of underlying data and models. Particularly, with older studies, simply finding the data sets and determining ownership may be expensive or impossible. For older studies with human subjects, obtaining consent to release of data may be practically impossible,

---

[150] *See,* 82 Fed. Reg. 7,149-7,274.
[151] *Id.*
[152] HHS, *Federalwide Assurance (FWA) for the Protection of Human Subjects*, https://www.hhs.gov/ohrp/register-irbs-and-obtain-fwas/fwas/fwa-protection-of-human-subjecct/index.html (last accessed Aug. 13, 2018).
[153] Memorandum by Alison Cullen, Chair, SAB Work Group on EPA Planned Actions for SAB Consideration of the Underlying Science (May 12, 2018).

and the data may have been collected in ways that would make protecting privacy with release difficult or impossible.[154]

For some studies, administrative issues related to the data could be the most difficult barrier to overcome in providing for public release. Larger and more costly studies are often performed by groups of researchers within a university, across multiple institutions, or across multiple individual companies. Over time, the data itself may become lost or misplaced, or it may become unclear who actually owns and controls access to the data. Academics move among institutions, companies merge and spin off, and the initial agreements were not always clear in the first instance. Obtaining consent from multiple institutional players takes extensive time and resources, at minimum, and simply may no longer be possible in some instances.[155]

These problems are exacerbated with respect to human subject studies. Researchers are legally and ethically obliged either to protect the privacy of the individual study subjects or attain each subject's consent to share data.[156] This can be impractical for older studies and virtually impossible for larger studies, and extremely burdensome. For example, the Harvard Six cities study was started in 1975 and had 8,111 participants.[157] The ACS CPSII extended analysis by Krewski in 2009, which is central to $PM_{2.5}$ NAAQS standards, was initiated in 1979 and encompassed data from 500,000 study participants who lived in 116 metropolitan areas.[158] For these types of situations, tracking down participants (or where the participants have passed away, their family members) to get consent is simply not realistically possible.

Even in situations where investigators might theoretically be able to attain consent, it would require extensive financial and human resources, which are usually simply not available, especially to academic researchers or to EPA. EPA ignores this prohibitive constraint and makes no attempt to address it.

c) There are additional significant barriers to public release of data in some situations, even for prospective studies.

Even with respect to prospective application of EPA's proposal, providing for public release of underlying data and models is costly and resource intensive, creating a serious disincentive for researchers to meet EPA's proposed requirements. Investigators willing to make their study underlying data publicly available would still face the logistical hurdle of making the data and models available in a manner sufficient for independent validation by the public. In

---

[154] *See, e.g.,* National Academies of Sciences, Engineering, and Medicine, *Principles and obstacles for sharing data from environmental health research: Workshop summary*, 61-63 The National Academies Press (2016), https://www.nap.edu/catalog/21703/principles-and-obstacles-for-sharing-data-from-environmental-health-research.

[155] *Id.* at 45.

[156] *Federal Policy for the Protection of Human Subjects; Final Rule*, 82 Fed. Reg. 7,149 (Jan. 19, 2017); HIPAA Privacy Rule, 45 C.F.R. §§ 160, 164.102-106, 164.500-534.

[157] Dockery, D.W., Pope, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris Jr, B.G. and Speizer, F.E., *An association between air pollution and mortality in six US cities,* 329(24) New England Journal of Medicine, 1753-1759 (1993).

[158] Krewski D, Jerrett M, Burnett RT, et al., *Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality*, 140 Health Effects Institute, Boston MA (2009).

addition to the cost of thoughtful and effective deidentification or redaction of sensitive information, the proposed text would likely require researchers to prepare annotated manuals including precise detail as to what variables were collected, how information was collected, and the rationale for each step taken. Some manuals alone run into hundreds of pages. One press account noted the example of publicly available datasets from the National Center for Health Statistics, which can come with 100-page manuals; researchers would need to hire additional staff to meet such requirements.[159] Yet EPA fails even to recognize (much less propose any means to address) the cost to researchers in time and money, on top of the constraints on academic research already imposed by the very limited funding available for this type of work.

In addition, there are other barriers to public release of underlying data. Studies conducted on behalf of industry or with industry cooperation may contain confidential business information, the release of which could jeopardize a company's competitiveness.

Also, in some instances, researchers cannot make their data sets public without losing much of the value to the researcher of these laboriously and meticulously collected sets of information. Research, especially those studies that include large numbers of human subjects, are incredibly human and capital intensive endeavors. Moreover researchers may base years of work and multiple papers on unique datasets they developed and hold, and many scientists build their careers on carefully harvesting information from single large studies for years to come. It is not only unreasonable, but also unfair, to expect academic scientists to turn over their intellectual property and research investments, forgoing potential earnings and career advancements. Moreover, EPA's myopic and inflexible approach to data access gives no consideration to data sharing arrangements between researchers and the agency that could be developed to support EPA's consideration and integration of research.

If scientists are forced to choose between giving away their hard-earned data or forgoing any regulatory impact, it will discourage scientists from engaging in critical science that is targeted to help prevent disease and disability in our population. It appears that in many cases, scientists will choose to retain their datasets, with a worst-of-both-worlds result—EPA will be deprived of valid scientific information and the scientific community will be discouraged from contributing their critical expertise to policy-making. EPA's Proposal does not consider the real-world implications of forcing such choices on researchers.

The agency's failure to consider or examine any of these legitimate reasons for not making data publicly available is arbitrary and capricious.

2. The Proposal fails to propose any actual solutions to remedy the legitimate reasons for why data may not be made publicly available.

In the proposal EPA blithely and irrationally ignores or assumes away the real and significant issues raised above, suggesting that existing mechanisms and techniques can be used

[159] Alessandra Potenza and Rachel Becker, *Scott Pruitt's new 'secret science' proposal is the wrong way to increase transparency. Here's what scientists think a science transparency rule should include*, The Verge (May 1, 2018, 8:30am EDT), https://www.theverge.com/2018/5/1/17304298/epa-science-transparency-rule-scott-pruitt-data-sharing.

to protect privacy and confidentiality while making underlying research data publicly available. In fact, the evidence (including several of the sources that EPA cites) indicates that the potential mechanisms alluded to by EPA would only have the potential to address some of the barriers cited above, have serious limitations even for those, and are actually becoming less effective as it becomes easier to combine and manipulate public data sets.

a) EPA vaguely references a range of possible approaches to protecting privacy and confidentiality, but provides no evidence that any of these are sufficient to address the legitimate concerns raised above.

EPA vaguely claims "concerns about access to confidential or private information can, in many cases, be addressed through the application of solutions commonly in use across some parts of the Federal government."[160] EPA claims that there are examples from the Department of Health and Human Services, the National Institute of Standards and Technology, the Department of Education, and the Census Bureau. Unfortunately, apart from a reference to HHS guidance on data de-identification (discussed below), EPA does not actually identify or cite to any specific examples from these agencies in the proposed rule itself, making it impossible to discern what examples EPA believes exist or to meaningfully comment upon the degree to which such examples, if they exist, might suggest that these issues are manageable. The additional hyperlinks added to the docket on May 25, 2018, weeks into the comment period, also link to examples that provide no further assurance that this proposal can be implemented without implicating privacy concerns, and as discussed in detail below, the vaguely referenced other agencies' "solutions" are unlikely to be of much help.

The "solutions" EPA might have in mind do not address the issues raised by the Proposal because no other agency has tried to implement a requirement such as the one EPA proposes. Other agencies provide guidance and techniques to protect privacy during data collection and disclosure to allow more use of data collected by the *government*, not to mandate that data collected by academic or industry researchers be publicly available for purposes of replicating analyses. The Department of Education, for example, has shared techniques for institutions to provide data on students and schools to meet reporting requirements without compromising privacy.[161] They recognize that each technique "requires some loss of information."[162] While de-identified information may still be useful, e.g., to show overall school progress, in the context of the Education Department, it is not clear these techniques are transferable to other contexts.

EPA links to a document of the Privacy Technical Assistance Center, *Data De-identification: An Overview of Basic Terms*, which provides a high-level overview of key terms and practices to help educational agencies and institutions comply with the Family Educational Rights and Privacy Act (FERPA).[163] This document is concerned with data disclosure that occurs

---

[160] 83 Fed. Reg. 18,770.

[161] National Center for Education Statistics, *SLDS Technical Brief: Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting* (Dec. 2010), https://nces.ed.gov/pubs2011/2011603.pdf.

[162] *Id.* at 27.

[163] Privacy Technical Assistance Center, *Data De-identification: An Overview of Basic Terms* (2001), https://studentprivacy.ed.gov/sites/default/files/resourcedocument/file/datadeidentificationterms.pdf.

"when schools, districts, or states publish reports on student achievement or share students' data with external researchers" not to make underlying data publicly available for independent validation.[164] Thus, it is unclear that methods used to de-identify but preserve data for those purposes would be adequate in this context. For example, one of the methods that the U.S. Department of Education uses for disclosure avoidance for tabular data is to not release information for any cell that has a size below some minimum, which essentially means not disclosing information where there are small numbers in a certain cell.[165] Thus, it is quite possible that techniques that result in a loss of information would prevent researchers from repeating the experiment. Yet EPA fails to acknowledge the nuances and limitations of these policies.

EPA links to a NIST document entitled *De-Identification of Personal Information* by Simson L. Garfinkel (NISTIR 8053), which discusses de-identification, but not in the context of making research data publicly available for independently validating scientific studies. The document instead notes that "that there is a trade-off between the amount of de-identification and the utility of the resulting data" and that "[i]t is thus the role of the data controller, standards bodies, regulators, lawmakers and courts to determine the appropriate level of security, and thereby the acceptable trade-off between de-identification and utility."[166] It further notes that "de-identification approaches based on suppressing or generalizing specific fields in a database cannot provide absolute privacy guarantees, because there is always a chance that the remaining data can be re-identified using an auxiliary dataset."[167]

EPA's reference to the U.S. Census Bureau is similarly unhelpful. Here EPA provides a link to a website titled *Data Ingest and Linkage* that details the U.S. Census Bureau's approach to linking data across many records they hold.[168] The Website links to a working paper that describes the method by which the Census assigns a unique person identifier to records it holds that enables it to link records together to create the final file.[169] It is totally unclear how this process on linking together records is a solution that EPA could implement to protect privacy of individuals when disclosing data as it concerns how to identify data with specific people—not protecting privacy.

While other agencies are clearly grappling with the issue of how to make government-collected data available, they have also highlighted the many challenges in protecting privacy and confidentiality while doing so—such as the ability for de-identified data to be re-identified—and these agencies accept that there is more work to be done before these concerns are fully

---

[164] *Id.* at 1.
[165] *Id.* at 4.
[166] Simson L. Garfinkel, *De-Identification of Personal Information* (NISTIR 8053), 11-12 NIST (Oct. 2015), https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf.
[167] *Id.* at 5.
[168] U.S. Census Bureau, *Data Ingest and Linkage*, https://www.census.gov/about/adrm/linkage/technical-documentation/processing-de-identification.html (last accessed Aug. 13, 2015).
[169] Deborah Wagner & Mary Layne, *The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software*, CARRA Working Paper Series, Working Paper # 2014-01, U.S. Census Bureau (July 1, 2014).

addressed.[170] The letter filed in this docket by the Presidents of the National Academies of Science, Engineering and Medicine underscores these difficulties, specifically noting the National Academies' previous work finding that "statistical analyses of data sets that generate highly precise results—such as geographic specificity or other characteristics that identify respondents—may result in privacy breaches . . . This presents a new challenge that federal statistical agencies are just beginning to address."[171] EPA does not even acknowledge, much less try to address, these gaps in agencies' abilities to protect sensitive data.

EPA cursorily mentions a range of options for facilitating secure access to confidential data, including: "[r]equiring applications for access; restricting access to data for the purposes of replication, validation, and sensitivity evaluation; establishing physical controls on data storage; online training for researchers; and nondisclosure agreements."[172] EPA does not indicate whether it would deem providing access with these types of controls in place sufficient to meet EPA's proposed requirement "publicly available in a manner sufficient for independent validation." EPA also fails to recognize the significant costs associated with implementing most of these options or the risks to privacy that remain even if these methods are employed.

b) EPA cites to one example—the technique of deidentification—but fails to acknowledge, let alone address, the significant costs and limitations of this approach.

As already discussed, it is legally and ethically necessary to ensure the privacy of the individuals whose data have been collected, as some of these data, such as medical history or employment data, can be quite sensitive. EPA suggests deidentification and redaction of sensitive information can be used to protect privacy when study data is made public. EPA fails to recognize that these techniques are generally burdensome and costly, and may lose too much information for replication purposes. EPA also ignores the real concerns, based in empirical evidence, about reidentification of individuals through cross linking with existing public datasets and the ensuing breach of privacy.[173]

---

[170] *See, e.g.*, Simson L. Garfinkel, *De-Identification of Personal Information* (NISTIR 8053), NIST (Oct. 2015) (detailing methods of re-identification and challenges to de-identifying information, concluding "there is comparatively little known about the underlying science of de-identification" and "there is a clear need for standards and assessment techniques that can measurably address the breadth of data and risks described in this paper.").

[171] Letter to Acting Administrator Wheeler from Marcia McNutt, President of the National Academy of Sciences, C.D. Mote, Jr., President of the National Academy of Engineering, and Victor J. Dzau, President of the National Academy of Medicine at 4 (July 16, 2018) (citations removed).

[172] 83 Fed. Reg. 18,771.

[173] "Recently, a peer reviewed study examined the identifiability of records from an environmental health study in Northern California. Using data considered by HIPAA to be sufficiently de-identified to be made public, which involved far fewer variables than would be required to make public in the cohort studies, they were able to correctly identify over 25% of the participants. Another study searched the Lexis-Nexis database for stories that mentioned hospitalization, and by matching that with age, race, sex and Zip code from a supposedly anonymized hospital admissions data base was able to match 43% of the people named in the news stories to their medical records." Comments of the International Society for Environmental Epidemiology on EPA's proposed rule on Strengthening Transparency in Regulatory Science (EPA-HQ-OA2018-0259-0001), https://www.regulations.gov/document?D=EPA-HQ-OA-2018-0259-1973 (citing Sweeney L, Yoo JS, Perovich L, Boronow KE, Brown P and JG B., *Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one*

Indeed, experts have observed that even the disclosure of redacted or "de-identified" data sets has become more fraught as public health studies have become more rigorous, because these studies are relying upon greater quantities of ever more granular personal information.[174]

i. Deidentification is complicated and costly.

EPA states that "[o]ther federal agencies have developed tools and methods to deidentify private information," but then cites to only one source, which does not address the concerns raised here.[175] EPA cites to the U.S. Department of Health and Human Services' *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*.[176] This guidance provides two methods for de-identifying data: (1) expert determination method, where an expert determines that, after application of statistical and scientific principals and methods, the risk is very small that the information alone or with other available information could be used to identify the subject; and (2) the safe harbor method, requiring that a number of identifiers are removed.[177] The first method requires case-by-case work, and EPA has provided no information regarding how EPA or others could potentially implement it or how much it might cost. In addition, there is no indication of how broadly this technique might be applicable to adequately de-identify data. *I.e.*, EPA must provide its views on whether this technique is likely to be applicable to the majority of studies relevant to EPA with non-public data, some studies, or only a handful. The second method requires removal of much information that may be necessary to be able to reanalyze or reproduce the research results, so it is unclear whether it would satisfy EPA's requirements in the Proposal. The second method is also costly, which EPA also completely disregards. Furthermore, even the safe harbor method has been shown to provide potentially insufficient privacy protections due to the mosaic effect, discussed more below.

EPA further states: "The National Academies have noted that simple data masking, coding, and de-identification techniques have been developed over the last half century. . . ," seemingly suggesting that data can easily be modified to address privacy concerns.[178] This is incorrect. The National Academies in fact recognizes that complex, evolving, and yet undeveloped techniques are needed to resolve these concerns: "Initially, relatively simple data masking techniques, such as top coding income amounts. . .were used to generate restricted data

---

environmental health study, Technology Science (2017) and Sweeney L., *Only You, Your Doctor, and Many Others May Know*, Technology Science (2015)).

[174] *See* Letter from Daniel S. Greenbaum, Health Effects Institute, to Lek Kadeli, Environmental Protection Agency 3 (Aug. 27, 2013) (describing the use of increasingly fine-grained community-level and zip code-level data in public health studies, and noting that "these characteristics – which have in general enhanced the quality and the sensitivity of the studies – increase the difficulty of providing a fully "de-identified" data set while also enabling a different investigator to conduct a full replication and sensitivity analysis of the original study results.").

[175] 83 Fed. Reg. at 18,771.

[176] 83 Fed. Reg. at 18,771 n. 17.

[177] HHS, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

[178] 83 Fed. Reg. at 18,771.

products [,] [d]uring the last decade the increasing risks of confidentiality breaches have led researchers to develop increasingly sophisticated methodologies for restricted data products."[179] They state, "more research is clearly needed to assess the relative ability of different masking methods, and of synthetic data, to reduce the risk of disclosure while preserving data utility."[180] They recognize the current limitations of producing restricted data that sufficiently limits identifiability to allow it to be made publicly available in a useful form. They note that "well-informed policy making" requires "[r]esearch using detailed confidential data" that cannot be made public—which the Proposal fails to acknowledge to the detriment of the quality of EPA's policy decisions.[181] In the meantime, the National Academies state that more work is needed to allow "[h]igh-quality public-use files" that still assure "the inferential validity of the data while safeguarding their confidentiality."[182]

ii. Ongoing developments in data analytics make data deidentification more difficult to conduct and less likely to adequately protect privacy and confidentiality.

In pointing to the option of deidentification and redaction techniques, EPA also fails even to mention, let alone address, the increasing risk of re-identification through data analysis using multiple data sets. The so-called "mosaic effect" makes even very limited, redacted releases of data to the public a threat to the privacy of study subjects. OMB has recognized the threat to privacy from the mosaic effect, which it describes as "when the information in an individual dataset, in isolation, may not pose a risk of identifying an individual (or threatening some other important interest such as security), but when combined with other available information, could pose such risk."[183] OMB specifically highlighted the complicated nature of this threat and the need for agencies to address it carefully, particularly as they may not possess the needed expertise.[184]

Studies show the reality and scope of the re-identification threat. For example, Dr. Latanya Sweeney, professor of government and technology in residence at Harvard University, has examined deidentified datasets and combined them with other public data sets to test this concern. She was able to use information in medical information and a voter list, such as birth date, gender, and zip code, to identify individuals in the deidentified Massachusetts Group Health Insurance Commission dataset in 1997, including the then Massachusetts Governor,

---

[179] National Research Council, *Expanding Access to Research Data: Reconciling Risks and Opportunities*, 27 The National Academies Press (2005).
[180] *Id.* at 28.
[181] *Id.* at 2.
[182] *Id.*
[183] OMB Memorandum M-13-13, Memorandum for the Heads of Executive Departments and Agencies on Open Data Policy—Managing Information as an Asset 4-5 (May 9, 2013).
[184] *Id.* at 9-10 ("Agencies should note that the mosaic effect demands a risk-based analysis, often utilizing statistical methods whose parameters can change over time, depending on the nature of the information, the availability of other information, and the technology in place that could facilitate the process of identification. Because of the complexity of this analysis and the scope of data involved, agencies may choose to take advantage of entities in the Executive Branch that may have relevant expertise, including the staff of Data.gov.")

William Weld.[185] Studies have indicated that between 63% and 87% of the population of the United States could be uniquely identified by using only gender, ZIP code, and date of birth.[186] Dr. Sweeney was also able to link data in the Personal Genome Project to names and contact information, identifying between 84 to 97% of profiles.[187] In 2011 she was able to identify 43% of individuals in a department of health in Washington state hospital discharge database using newspaper stories.[188] Another study[189] showed how "data on air and dust samples from 50 homes in two communities in California could be combined with data released under the Safe Harbor provisions of the Health Insurance Portability and Accountability Act (HIPAA) to 'uniquely and correctly identify [in one community] 8 of 32 (25 percent) by name and 9 of 32 (28 percent) by address.'"[190]

The Commission on Evidence-Based Policymaking, which EPA also cites in the Proposal[191], also stresses the dangers of re-identification of data that has been stripped of direct identifiers. They note: "No existing statistical disclosure limitation method. . . is able to completely eliminate the risk of re-identification," despite increasingly complex techniques that have been developed since the 1970s.[192] They also note the threat posed by the "cumulative amount of information available about individuals and businesses that could be used for re-identification,"[193] with the threat increasing as available information grows and technology to allow re-identification improves.[194]

Further, the National Academies note, "data that are most useful to legitimate researchers typically have characteristics that pose substantial risk of disclosure."[195] This includes information such as:

- detailed geographic information;
- repeated data collection from the same subjects;
- outliers, such as people with very high incomes;
- many attribute variables; and

---

[185] Rothstein, Mark A., *Is deidentification sufficient to protect health privacy in research?*, 10.9 The American Journal of Bioethics 3-11, 6 (2010).
[186] *Id.* at 5.
[187] Sweeney, Latanya and Abu, Akua and Winn, Julia, *Identifying Participants in the Personal Genome Project by Name* (April 29, 2013), https://ssrn.com/abstract=2257732 or http://dx.doi.org/10.2139/ssrn.2257732.
[188] Sweeney L., *Matching known patients to health records in Washington State data*, Harvard University, Data Privacy Lab (2013), https://dataprivacylab.org/projects/wa/1089-1.pdf.
[189] Latanya Sweeney, Ji Su Yon, Laura Perovich, Katherine E Boronow, Phil Brown, and Julia Green Brody, *Re-identification Risks in HIPAA Safe Harbor Data: A Study of Data From One Environmental Health Study*, Technology Science (Aug. 28, 2017), https://techscience.org/a/2017082801/.
[190] Commission on Evidence-Based Policymaking, *The Promise of Evidence-Based Policymaking*, 54 (2017), https://www.cep.gov/content/dam/cep/report/cep-final-report.pdf.
[191] 83 Fed. Reg. at 18771, n. 19.
[192] Commission on Evidence-Based Policymaking, *The Promise of Evidence-Based Policymaking* 53 (2017).
[193] *Id.* at 54.
[194] *Id.* at 55.
[195] National Research Council, *Expanding Access to Research Data: Reconciling Risks and Opportunities*, 21 The National Academies Press (2005).

- complete census data rather than a survey of a small sample of the population.[196]

There is increased vulnerability in "[d]ata with geographic detail, such as census block data" and longitudinal data obtained in panel surveys, which is often salient in environmental research.[197]

iii. Deidentification may make data sets unusable for reanalysis purposes.

Work by other experts in this area suggests that deidentification can be carried out and help protect privacy, but it may produce datasets that have lost vital information needed for specific analyses.[198] Even the HIPPAA guidelines document states: "Of course, de-identification leads to information loss which may limit the usefulness of the resulting health information."[199] Such results limit the utility of deidentified data sets and would not meet the requirements of the proposed rule which state that "*EPA will ensure that the data and models underlying the science is publicly available in a manner sufficient for validation and analysis*."

Further, even if it may be technically possible to release some amount of data while preserving privacy in some cases, doing so imposes substantial additional costs.[200] The preamble of the proposed rule suggests that privacy concerns can be addressed through mechanisms such as data masking, coding, and de-identification techniques—all of which would impose additional costs on researchers. The preamble also indicates that requirements for dose response data and availability may differ and involve a range of mechanisms such as deposition in public data repositories, and controlled access in federal research data centers—which would require EPA funding to maintain the facilities.[201] As discussed further in Section V of these comments, the proposed rule fails to acknowledge these costs, let alone provide any information about them or suggest ways to provide for them. Nevertheless, the costs can be significant, and even smaller costs could be prohibitive for many researchers.

At a time when federal funding for research in environmental and public health-related fields has largely flat-lined, academic researchers, in particular, are likely to have few additional

---

[196] *Id.* at 21-22.
[197] *Id.* at 22.
[198] Simson L. Garfinkel, *De-Identification of Personal Information* (NISTIR 8053), NIST (Oct. 2015) (saying the goals of allowing data to be used while providing privacy protections "are antagonistic, in that there is a trade-off between the amount of de-identification and the utility of the resulting data.").
[199] HHS, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html.
[200] National Academies of Sciences, Engineering, and Medicine, *Principles and obstacles for sharing data from environmental health research: Workshop summary*, 46-47 The National Academies Press (2016), https://www.nap.edu/catalog/21703/principles-and-obstacles-for-sharing-data-from-environmental-health-research.
[201] *See,* The National Academies, *Improving Access to and Confidentiality of Research Data: Report of a Workshop*, National Academies Press 48 (2000) (At present, [costs for federal research data centers] are being covered partly by federal agency budgets and partly by user fees. The Census Bureau's research data centers have been supported in part by grants from the National Science Foundation and NIA, but may eventually have to recover more of their costs from users.").

funds available to undertake these activities.[202] This raises additional concerns—if researchers funded by industry are generally able to support the additional costs of making data publicly available, while academic researchers are far less likely to be able to do so, EPA's proposed approach could institutionalize a dangerous bias in the source of studies that EPA is allowed to use for regulatory activity.

With respect to the potentially very large costs that would accrue to EPA, EPA's proposal provides no indication that any funding to support such activities would be available. EPA funding is at its lowest level since the 1980s.[203] Absent a significant change in Congressional priorities, any EPA expenditures for the purposes of supporting making data publicly available would necessarily require cutbacks in other critical areas of environmental protection, which might include supporting additional research, conducting inspections, issuing permits, setting standards, or many other activities. EPA's Proposal includes no discussion of whether funds would be made available, nor whether other activities would be sacrificed, whether these trade-offs would make any sense, and what the overall impacts might be on public health and the environment.

B. The Proposal Will Not Advance the Supposed Cause of "Transparency" Upon Which it is Based.

The Proposal does not present or support the case that public accessibility to underlying data is necessary to vet scientific research—which, as discussed above, it is not—but even if it was, as discussed above, the scientific community is already taking steps to make underlying data publicly available where feasible, with the widespread understanding that this is neither necessary nor appropriate in all cases.[204] The Proposal does not examine the policies and practices that are already working to make data publicly available where feasible, the extent to which existing policies may already be sufficient to meet EPA's alleged transparency goals, or the reasons why some data is still not released publicly. Still less does EPA question whether this proposal would add anything to the current efforts, or whether it would have any effect whatsoever in increasing public accessibility of data.

1. Where there are lower hurdles to making data publicly available, this is already commonly occurring, with support from various initiatives.

---

[202] *See*, American Association for the Advancement of Science, *Trends in Federal Research by Discipline FY 1970-2017*, chart, (last updated July 2018), http://mcmprodaaas.s3.amazonaws.com/s3fs-public/Disc-1_0.jpg?RrBDGaSpG5edeDsiBRyoQyApdamjOs4O.

[203] *Compare* FY 2018 budget of $5.655 billion (EPA, *FY 2018 Budget in Brief* (May 2017)) and projected FY 2019 EPA budget of $6.146 billion (EPA News Release, *EPA FY 2019 Budget Proposal Released* (Feb. 12, 2018), https://www.epa.gov/newsreleases/epa-fy-2019-budget-proposal-released) with fiscal year 2017's budget of $8.058 billion and historical budgets (*EPA's Budget and Spending*, https://www.epa.gov/planandbudget/budget (last accessed July 26, 2018)).

[204] *See* National Academies of Sciences, Engineering, and Medicine, *Principles and obstacles for sharing data from environmental health research: Workshop summary*, The National Academies Press (2016), https://www.nap.edu/catalog/21703/principles-and-obstacles-for-sharing-data-from-environmental-health-research.

There are already various ongoing initiatives to make scientific data and models more commonly publicly available, where appropriate, as discussed more below. For example, EPA cites the ongoing implementation of the 2016 *Plan to Increase Access to Results of EPA-Funded Scientific Research*.[205] This Plan aims to maximize access to "research data underlying a publication" resulting from EPA-funded research.[206] It is worth emphasizing the Plan also exempts "research data [that] cannot be released due to one or more of constraints, such as requirements to protect confidentiality, personal privacy, proprietary interest, or property rights."[207] There is also a 12-month embargo period before publications are made publicly available.[208] The Plan also explicitly indicates that

> [i]t is important to recognize that some research data cannot be made fully available to the public but instead may need to be made available in more limited ways, e.g., establishing data use agreements with researchers that respect necessary protections. *Whether research data are fully available to the public or available to researchers through other means does not affect the validity of the scientific conclusions from peer-reviewed research publications.*[209]

EPA also mentions the data availability policies or requirements of many scientific journals (although EPA does not specifically discuss any of these policies or indicate how or why they are not sufficient to address EPA's concerns).[210] Thus, where there are not significant barriers due to costs, or confidentiality or other concerns, there are increasing mechanisms to encourage scientists to make their data meaningfully and responsibly publicly available, and in response to these mechanisms, scientists frequently do so already.[211]

2. EPA's proposed approach does not require researchers to make underlying data publicly available.

There are multiple real and significant barriers to the public release of underlying data from some studies, and the Proposal cites no reason to believe that, in the majority of cases where data is not already released, one or more of those barriers are not present. Because those barriers are significant, this is not a situation where creating an incentive to private action is likely to be sufficient to drive such action where it is not already occurring.

---

[205] 83 Fed. Reg. at 18770.
[206] EPA, *Plan to Increase Access to Results of EPA-Funded Scientific Research* 11 (Nov. 29, 2016), https://www.epa.gov/sites/production/files/2016-12/documents/epascientificresearchtransperancyplan.pdf.
[207] *Id.* at 11.
[208] *Id.*
[209] *Id.* at 4-5 (emphasis added).
[210] 83 Fed. Reg. at 18,770 (stating that the policies and recommendations EPA considered were "informed by the policies recently adopted by some major scientific journals and cites to "related policies from the Proceedings of the National Academy of Sciences, PLOS ONE, Science, and Nature."); 83 Fed. Reg. at 18,771 n. 20 (claiming the "policies or recommendations of publishers Taylor & Francis, Elsevier, PLOS, and Springer Nature" support the Proposal because they require authors to deposit the data underlying their studies in public data repositories).
[211] Jeremy Berg, *Obfuscating with transparency*, 360 Science 133 (Apr. 13, 2018), http://science.sciencemag.org/content/360/6385/133/tab-pdf ("Increasingly, many publications, including those from the Science family of journals, are linked to underlying data in accessible forms in repositories where they are readily available to interested parties, particularly those who seek to reproduce results or extend the analysis.").

Yet, with respect to release of data, the Proposal would only create an incentive for private action, not an actual requirement that data be released. First, this Proposal addresses data produced and held by external scientists, not data held by EPA itself or that EPA has authority to gain access to. Where EPA holds data, it is already governed by the Information Quality Act, OMB Circular A-110, and the Freedom of Information Act.[212] The Shelby Amendment required OMB to amend Circular A-110 to require that federal agencies provide "research data relating to published research findings produced under an award that were used by the Federal Government in developing an agency action that has the force and effect of law" to the public through the Freedom of Information Act.[213] Importantly, the term "research data" excludes "[t]rade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law" as well as "[p]ersonnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study."[214] Many voiced concerns that even this provision could compromise scientific research and personal privacy.[215] This Proposal presumably is also not directed at studies funded by EPA, where the researchers must generally make data publicly available as a condition of receiving funding.[216] There are already mechanisms by which EPA is making research data publicly available where it has the authority and access to do so, and only after carefully ensuring that doing so will not compromise privacy interests.

Second, EPA has no authority to regulate the authors of studies or the scientific journals in which the studies are published, and EPA makes no attempt to regulate them directly. The preamble to the proposed rule states: "EPA should ensure that the data and models underlying scientific studies that are pivotal to the regulatory action are available to the public."[217] It further states that the proposed regulation is "designed to provide a mechanism to increase access to dose response data and models underlying pivotal regulatory science...."[218] The proposed regulations then state that for significant regulatory actions EPA "shall ensure that dose response data and models underlying pivotal regulatory science are publicly available in a manner

---

[212] OMB Circular A-110 Revised 11/19/93 As Further Amended 9/30/99 36(d)(1) ("In addition, in response to a Freedom of Information Act (FOIA) request for research data relating to published research findings produced under an award that were used by the Federal Government in developing an agency action that has the force and effect of law, the Federal awarding agency shall request, and the recipient shall provide, within a reasonable time, the research data so that they can be made available to the public through the procedures established under the FOIA."); *See also*, Lynn R. Goldman & Ellen K Silbergeld, *Assuring Access to Data for Chemical Evaluation*, 121 Environmental Health Perspectives 149 (Feb. 2013), https://ehp.niehs.nih.gov/wp-content/uploads/121/2/ehp.1206101.pdf (noting the numerous feasibility concerns that would arise were EPA to be required to make raw underlying data available for studies not governed by these mechanisms , given the large number of studies it usually relies on and that fact that EPA is usually not in possession of the raw data, in addition to funding and ethical limitations).

[213] OMB Circular A-110 (36)(d)(1).

[214] OMB Circular A-110 (36)(d)(2)(i).

[215] *See* Eric A. Fischer, *Public Access to Data from Federally Funded Research: Provisions in OMB Circular A-110*, Congressional Research Service, 13 (Mar. 1, 2013), https://fas.org/sgp/crs/secrecy/R42983.pdf.

[216] U.S. EPA, *Plan to Increase Access to Results of EPA-Funded Scientific Research* (Nov. 29, 2016), https://www.epa.gov/sites/production/files/2016-12/documents/epascientificresearchtransperancyplan.pdf.

[217] 83 Fed. Reg. at 18769.

[218] 83 Fed. Reg. at 18770.

sufficient for independent validation." [219] But (apart from studies that EPA funds) EPA has no authority to require those data and models to be made public.

Hence, this proposal would regulate not the scientists, but EPA itself. EPA would "ensure" that data and models underlying scientific studies "pivotal" to regulatory action are publicly available *simply by barring EPA's own use in regulatory actions of any studies for which the authors do not make the data and models publicly available.* The "mechanism" mentioned in the preamble is not technical assistance or funding to encourage greater availability of data; it is simply the pressure generated by EPA's refusal to consider the results of a study if the authors do not release publicly the underlying data and models. The obvious question that EPA has neither asked nor attempted to answer in the Proposal is whether such a ban would be sufficient to incentivize study authors to make their data and models publicly available, where they have not already done so, or whether the ban will largely result in just limiting the studies available to EPA. Most of the significant barriers to release detailed above are not a matter of the researcher's preference, but rather take the form of legal and ethical constraints, significant costs, large time investments, or the loss of proprietary data critical to a researcher's future career prospects. While it seems plausible that having their research applied in a regulatory context would be viewed as an incentive by some, or perhaps many, researchers, there is no reason to believe that such an incentive would be sufficient to overcome the significant barriers to public release of data where those barriers exist. Indeed, the party most likely to be incentivized by EPA's proposed requirements is the regulated community which has vested financial interests in regulatory actions the agency may take—a situation that almost certainly will lead to significant bias and conflicts of interests in the scientific evidence that the agency considers.

Yet EPA barely acknowledges the nature of the "mechanism" it is proposing, and EPA certainly does not explore in any way how the mechanism would operate or whether it would be effective in driving release of data. Still less does EPA admit that the primary effect of this approach is very likely to be the exclusion of critical valid scientific studies from EPA's consideration. Finally, EPA utterly fails to contemplate what the effect of such exclusion would be on EPA's ability to adopt regulatory standards that protect public health and the environment.

C. The Proposal does not Acknowledge, Much Less Examine, its Likely Actual Effect—Reducing the Quality and Quantity of Studies upon which Regulatory Decisions are Based.

1. EPA fails to recognize that forcing the disclosure of all data and models would have harmful effects on the quality and quantity of scientific research used by EPA.

Although it appears highly unlikely that this proposal would drive additional data to be released, EPA presumes otherwise, and fails to recognize the harms that would likely result if EPA actually were successful in finalizing the rule. One reason researchers are particularly cautious about releasing human subjects data is that they understand that public willingness to

[219] 83 Fed. Reg. 18773.

participate in research studies depends upon protecting the privacy of the participants. Risks of privacy breaches and researchers' inability to control use of subject data will undermine potential participants' confidence in scientists' ability to protect their information.[220] This will likely reduce participation in studies or even lead to biases in responses from participants.[221] It could also result in attrition of participation by select subpopulations, particularly those who may be most vulnerable, such as children or people with disabilities or disease, or those with the most to protect, such as high socioeconomic populations. Reduced participation and particularly reduced participation among select subpopulations will reduce scientists' ability to draw meaningful inferences from their results to broader populations, the whole of which EPA is charged with protecting.

In addition, the prospect that their research would not be used if researchers were unable to make their data public is likely to deter researchers from even engaging in environmental health research, particularly research involving human subjects.[222] Lynn Goldman and Ellen Silbergeld conclude that a requirement by EPA that researchers release raw data underlying studies reviewed for rulemakings on pesticides and chemicals "would not be tenable" and would in fact "have a chilling effect on the engagement of the global scientific community in research relevant to the protection of human health and the environment."[223] Overall, the result will be to diminish and undermine the strength of the scientific information available to EPA.

2. Because EPA will be barred from using many valid scientific studies with nonpublic data, the net effect of this proposal will be to harm, not strengthen, EPA's use of science in the regulatory process.

The most damaging aspect of EPA's proposal is that it will bar EPA from using many valid scientific studies that provide critically important information supporting regulatory standards and requirements. This will significantly harm, not strengthen, EPA's use of science in the regulatory process—especially since the public availability of data is neither necessary nor sufficient to ensure the validity of the studies EPA relies upon. It is clearly arbitrary and

---

[220] *See* Eugenia Economos, Farmworker Association of Florida, Testimony at EPA Public Hearing on Proposed Rule "Strengthening Transparency in Regulatory Science" (July 17, 2018); Leila Jamal et. al, *Research Participants' Attitudes Towards the Confidentiality of Genomic Sequence Information*, 22 Eur. J. Hum. Genetics 964 (2014), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4350593/.

[221] Christine Lothen-Kline et al., *Truth and Consequences: Ethics, Confidentiality, and Disclosure in Adolescent Longitudinal Prevention Research*, 33 Journal of Adolescent Health 385-394 (2003).

[222] *See* Augusta Wilson, Climate Sci. Legal Def. Fund, Testimony at EPA Public Hearing on Proposed Rule "Strengthening Transparency in Regulatory Science" (July 17, 2018), https://www.csldf.org/2018/07/16/why-we-oppose-to-the-epas-proposed-transparency-rule/ ("This could have a deeply concerning chilling effect on the conduct of important human health studies. Privacy concerns could influence what science gets done and what does not. Lines of scientific inquiry that would have been pursued may not be. The quality of data may be poorer than it otherwise would have been."); Augusta Wilson, *Big Tobacco's Smoke and Mirrors Revived by Pruitt's Science Transparency Policy*, The Hill (June 4, 2018, 5:00 PM), http://thehill.com/opinion/energy-environment/390638-big-tobaccos-smoke-and-mirrors-revived-by-pruitts-science ("Good scientists may understandably hesitate to pursue important lines of scientific inquiry if doing so will make them targets for regulators, interest groups and legislators who seek to impugn their credibility and troll through their emails looking for ways to publicly embarrass them.").

[223] Lynn R. Goldman & Ellen K Silbergeld, *Assuring Access to Data for Chemical Evaluation*, 121 Environmental Health Perspectives 149, 150 (Feb. 2013), https://ehp.niehs.nih.gov/wp-content/uploads/121/2/ehp.1206101.pdf.

capricious for EPA to sacrifice the agency's use of the best available science under these circumstances.

a) The prohibition on using studies with underlying nonpublic data will operate to exclude quality research results from EPA's regulatory process.

The next subsection provides an extensive discussion of some of the types of studies and specific studies that EPA would be unable to use under the Proposal.[224] Prior analyses by the Congressional Budget Office of related legislative proposals have also concluded that public availability requirements would significantly reduce the number of studies EPA relies upon—perhaps by as much as one-half.[225] Bizarrely, however, EPA does not even mention this probable effect of the Proposal, let alone provide information on which particular studies or types of studies would be excluded (absent a case-by-case exemption). Further, EPA utterly fails to consider what the effects of such exclusions could be on EPA's ability to develop and support standards to protect public health and the environment. There are many areas where these effects might be extremely damaging, as the examples below detail.

Not only would this proposal exclude valid studies, but it may well disproportionately exclude high quality studies. Some of the most robust and informative environmental health studies are human subjects studies with a large number of geographically distributed participants who are tracked over very long periods of time. These attributes make the results of these studies especially useful in regulatory decision making, since they are more representative of the population being addressed and provide information on exposure and health effects over a period of time. But these are also the attributes that make public release of the underlying data most difficult, and frequently impossible, as discussed above in Section II.A.1. Excluding these studies is highly likely to distort and undermine regulatory decision making by removing support for standards that are actually health protective. EPA has not identified any harms it is aiming to address through this Proposal, but whatever they are perceived to be, it is hard to see how they could outweigh the harm from barring EPA from considering the best available scientific information.

This Proposal also could be particularly harmful to EPA's ability to act in areas where the science is less developed, such as emerging threats. If there are a relatively small number of studies, the inability to consider some or all of them could cripple EPA's ability to act. This is

---

[224] Note that EPA has proposed to allow the Administrator to grant exemptions to the prohibition on a case-by-case basis, but the hurdle of requiring case-by-case determinations is so high (EPA relies on roughly 50,000 studies per year according to the CBO) and the criteria are sufficiently stringent (public availability must be "not feasible," which may well exclude, e.g., cost concerns) that it appears most plausible to assume that many studies will not be granted an exemption. *See* Section I.B.4 for further discussion.

[225] *See* Susanne S. Mehlman, Jon Sperl & Amy Petz, Cong. Budget Office, H.R. 1030: Secret Science Reform Act of 2015 at 2-3 (2015) ("CBO expects that EPA . . . would base its future work on fewer scientific studies . . . . CBO expects that the agency would probably cut the number of studies it relies on by about one-half . . . ."); Jon Sperl & Amy Petz, Cong. Budget Office, H.R. 1430: Honest and Open New EPA Science Treatment (HONEST) Act of 2017 at 1-2 (2017) ("EPA officials have explained to CBO that the agency would implement H.R. 1430 with minimal funding . . . . That approach to implementing the legislation would significantly reduce the number of studies that the agency relies on . . . .").

precisely the type of situation where a proactive early response could avoid extensive contamination (which is expensive to address) and multiple exposures (which are impossible to reverse), and the resulting adverse outcomes. Yet, apart from a question about how to apply the proposed rule to existing administrative records such as for the NAAQS, the closest EPA comes to hinting at the possibility of the regulatory and public health effects of excluding valid studies is when EPA asks the public to comment "on the effects of this proposed rule on individual EPA programs." None of these extremely consequential impacts of the Proposal are acknowledged or explored in any depth in the Proposal.

b) Examples of scientific studies that would be excluded

The proposed rule seeks to "ensure that dose response data and models underlying pivotal regulatory science are publicly available in a manner sufficient for independent validation."[226] The proposal indicates that "[i]nformation is considered 'publicly available in a manner sufficient for independent validation' when it includes the information necessary for the public to understand, assess, and replicate findings."[227] Further, footnote three of the proposal states:

> Historically, EPA has not consistently observed the policies underlying this proposal, and courts have at times upheld EPA's use [sic] non-public data in support of its regulatory actions. *See Coalition of Battery Recyclers Ass'n v. EPA, 604 F.3d 613, 623 (D.C. Cir. 2010); American Trucking Ass'ns v. EPA, 283 F.3d 355, 372 (D.C. Cir. 2002).* EPA is proposing to exercise its discretionary authority to establish a policy that would preclude it from using such data in future regulatory actions.[228]

Taken together, EPA is proposing to prohibit the use of studies involving dose response data and models in significant regulatory decisions where the underlying data are not publicly available. Such a prohibition would affect virtually all pending and future regulatory actions and, if applied retrospectively, past regulatory actions. Regulatory actions would not reflect the best available science, leading to inadequate or absent critical public health and environmental protections.

Eight examples of pending, past, and future regulatory actions that are themselves put at risk from the proposed regulation, or cite to studies that under the Proposal may not be able to be utilized in future actions, explained in more detail below, include:

- **proposed bans of trichloroethylene (TCE) for use in vapor degreasing, aerosol degreasing, and spot cleaning in dry cleaning facilities under TSCA section 6(a);[229]**

---

[226] 83 Fed. Reg. at 18773 (emphasis omitted).
[227] *Id.* at 18773–74.
[228] *Id.* at 18769 n.3.
[229] Trichloroethylene (TCE); Regulation of Use in Vapor Degreasing Under TSCA Section 6(a), 82 Fed. Reg. 7432 (Jan. 19, 2017); Trichloroethylene; Regulation of Certain Uses Under TSCA § 6(a), 81 Fed. Reg. 91,592 (Dec. 16, 2016).

- **proposed ban of methylene chloride for use in paint and coating removal under TSCA section 6(a);[230]**
- **final rule setting formaldehyde emission standards for composite wood products under TSCA Title VI;[231]**
- **National Primary Drinking Water Regulation for arsenic under the SDWA;[232]**
- **NAAQS for oxides of nitrogen under the CAA;[233]**
- **NAAQS for ozone under the CAA;[234]**
- **forthcoming proposed National Primary Drinking Water Regulation for perchlorate in development under the SDWA;[235] and**
- **future regulatory action on the perfluoroalkyl substances (PFASs) perfluorooctanoic acid (PFOA) and perfluorooctane sulfonate (PFOS) under SDWA and Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA).[236]**

Explanations of the likely effect of EPA's Proposal on these regulatory activities are described below.

Proposed bans of TCE for use in vapor degreasing, aerosol degreasing, and spot cleaning in dry cleaning facilities under TSCA section 6(a)

EPA has proposed two regulations under TSCA section 6(a) to ban the use of TCE in vapor degreasing, aerosol degreasing and spot cleaning in dry cleaning facilities.[237] Exposure to TCE is linked to several adverse health outcomes, including liver and kidney issues, developmental effects, and several forms of cancer.[238] The scientific basis for these proposed regulations is provided in the agency's 2014 risk assessment: *TSCA Work Plan Chemical Risk Assessment, Trichloroethylene: Degreasing, Spot Cleaning and Arts & Crafts Uses*[239] which

---

[230] Methylene Chloride and N-Methylpyrrolidone; Regulation of Certain Uses Under TSCA Section 6(e), 82 Fed. Reg. 7464 (Jan. 19, 2017).
[231] Formaldehyde Emission Standards for Composite Wood Products, 81 Fed. Reg. 89,674 (Dec. 12, 2016).
[232] National Primary Drinking Water Regulations; Arsenic and Clarifications to Compliance and New Source Contaminants Monitoring, 66 Fed. Reg. 6976 (Jan. 22, 2001).
[233] Review of the Primary National Ambient Air Quality Standards for Oxides of Nitrogen, 83 Fed. Reg. 17,226 (Apr. 18, 2018).
[234] National Ambient Air Quality Standards for Ozone, 80 Fed. Reg. 65,292 (Oct. 26, 2015).
[235] Drinking Water: Regulatory Determination on Perchlorate, 76 Fed. Reg. 7762 (Feb. 11, 2011).
[236] Press Release, EPA, In Case You Missed It: "EPA Chief Vows that Clean Drinking Water is National Priority" (May 22, 2018), https://www.epa.gov/newsreleases/case-you-missed-it-epa-chief-vows-clean-drinking-water-national-priority.
[237] 82 Fed. Reg. at 7432; 81 Fed. Reg. at 91,592
[238] 82 Fed. Reg. at 7435–36.
[239] EPA, Office of Chem. Safety & Pollution Prevention, EPA Doc. No. 740-R1-4002, "TSCA Work Plan Chemical Risk Assessment: Trichloroethylene: Degreasing, Spot Cleaning and Arts & Crafts Uses" (2014) [hereinafter TCE Work Plan Risk Assessment], https://www.epa.gov/sites/production/files/2014-11/documents/tce_opptworkplanchemra_final_062414.pdf.